

THIS WEEK

EDITORIALS

WORLD VIEW Not so fast, science is far from saved **p.133**



BUTTERFLIES British species flutter by earlier each year **p.134**

BARRIER GRIEF Australian floods could bring surge in coral-eating starfish **p.136**

Think big

The best way to manage national parks in the face of the effects of climate change is not to manage at the park level, but to work with landscapes. A new US initiative shows the way.

In 1882, the US conservationist George Bird Grinnell wrote about humans invading natural habitats as “the tide of immigration” that was then sweeping across the American West. “There is one spot left, a single rock about which this tide will break, and past which it will sweep, leaving it undefiled by the unsightly traces of civilization.” That rock was Yellowstone National Park, then just ten years old.

Thanks in large part to the success of Yellowstone, this rocks-in-the-tide or ‘protected area’ model has been adopted worldwide. Yellowstone remains the archetype for the park as an island in space and time, walled-off from changes to the land around it. But any park scientist or manager will tell you that to freeze a park in time is an unattainable ideal. And for better or worse, parks cannot be completely isolated in space either. Yellowstone is surrounded by national forests, ranches, game refuges and other natural lands that are ten times the size of the park itself, as well as by the spawning tendrils of residential development. European spotted knapweed gets in and grizzly bears get out.

As the effects of global climate change begin to unspool, park managers at Yellowstone and around the world are deciding how to proceed, torn between their impulse to fight to keep ecosystems the way they are and a reluctance to fiddle with nature too much (see page 150).

Perhaps the best approach is for them to ponder instead the larger landscape in which their parks sit. Scaling up is reassuring. At the park level, climate change may extirpate a species. At the landscape level, climate change merely moves it. And scaling up is more effective. Ecologists and conservation biologists have known for decades that small isolated parks leak species. Smaller populations have smaller gene pools in which maladapted traits are more likely to become fixed. Smaller populations are more vulnerable to drought, pests, hard winters or simple bad luck.

This is why conservation biologists, since at least the early 1990s, have called for parks to be connected to one another by unbroken corridors of nature, through which large species can move. For small mobile species, such as birds and insects, a stepping-stone scatter of protected areas close to one another has much the same effect. Climate change makes such connectivity even more important, as species challenged by the changing climate will need big gene pools to draw from and lots of different places to which they can move to. In particular, sites with microclimates to harbour species that can’t take the heat need to be identified, protected and linked to existing protected areas.

As corridor ecology has taken off as a scientific subfield, so have corridor and connectivity projects such as the Yellowstone to Yukon Conservation Initiative in North America and the Australian Alps to Atherton Connectivity Conservation Area. Britain’s Royal Society for the Protection of Birds is restoring lands beyond nature reserves in its Futurescapes programme. Some projects are even species-specific, such as Protecting the Pronghorn Path — a 240-kilometre-long federally designated and protected ‘migration corridor’ put in place on

all the different lands that the American antelope crosses on its way between summer and winter ranges in Wyoming. As the pronghorns make their way back and forth, the ungulates traipse across national forests, Bureau of Land Management gas fields, private cattle ranches and state-owned roads, where the department of transportation is this winter installing pronghorn-friendly underpasses. Coordinating

“It would be unforgivable to lose honeyeaters, antelopes, grizzlies and orchids.”

all of those players is a massive job, one that was tackled in this case by the Wildlife Conservation Society, based in New York. But there is not the money to do for the whole of Earth what the society was able to do in Wyoming.

In February 2010, the US Department of Interior ordered all the land-management agencies it oversees to join with other federal, state and private land managers in ‘landscape conservation cooperatives’ to help to understand and respond to the effects of climate change. At a recent scientific meeting in Yellowstone, many scientists groaned at the prospect of yet another entity in the already crowded and confusing realm of conservation planning. But if these cooperatives are widely embraced, they could be a way to move beyond the truism that landscape-level conservation is needed, and start to do it.

It would be unforgivable to lose honeyeaters, antelopes, grizzlies and orchids, not because scientists didn’t know how to save them, but because they were mired in bureaucratic mud. ■

Different strokes

Scientists in Romania and Bulgaria are having the best and the worst of times.

Marine geologist Liviu Giosan has lived through history. As a student in Romania he took part in the December 1989 demonstrations that brought down the communist government. Just months earlier, no one in the deeply isolated country would have believed that the hated dictatorship could ever fall. Yet the euphoria lasted barely six months. In June 1990, miners joined troops in violently crushing the street protests by students demonstrating against the communist presence in the newly elected government. After he got his degree in 1993, Giosan left for the United States to build a scientific career for himself — something he knew he couldn’t do at home. He is now an associate scientist at Woods Hole Oceanographic Institution, Massachusetts.

Most of the brightest young scientists in Romania — and

neighbouring Bulgaria, which shares a similar history — have emigrated as soon as they could. Very few have returned. It remains hard to do science in these countries, even though both joined the European Union in 2007. A few bright spots exist, but too much of the research landscape is still dominated by old-guard scientists who don't produce results, resist the introduction of international research standards and block the system to fresh blood.

As we report on page 142, the tide may just have turned for Romanian scientists. The government there is boosting funds and seems to know what is required for them to be spent wisely, and how to overcome scepticism among research émigrés. The Romanian government has a serious long-term plan for science, and this deserves recognition. Romanian scientists abroad, Giosan among them, are starting to smile.

The sentiment is not shared by those who watch the situation in Bulgaria with mounting despair. The Bulgarian government has only a short-term plan, the long-term consequences of which are likely to be disastrous. Funds have been slashed and the control of dozens of research institutes is set to be handed to the government from the Bulgarian Academy of Sciences, which will survive only as an academy. To separate active research from a learned society is not necessarily wrong — the status of both Britain's Royal Society and France's Académie Française demonstrates that — but the Bulgarian government is yet to show that it knows what to do with the institutes it is so keen to adopt.

In fact, it is clear that this populist government — which took office in July 2009 on an anti-corruption mandate — is not interested in science, and has convinced many among the general public that it is a waste of

money. Its science and education minister, Sergei Ignatov, was politically too weak to oppose a budget cut of more than one-third ordered in mid-2010. Science in Bulgaria has been humiliated as never before.

It is true that the Bulgarian Academy of Sciences needs deeper reform.

"A poorly performing science base cannot be fixed by just throwing money at it."

Under pressure from previous governments to raise its game, the academy organized an international evaluation and slimmed down to a fraction of the size it was in richer, Soviet times. But greater change is needed.

As the Romanian government has noted, a poorly performing science base cannot be fixed by just throwing money at it: regulations need to ensure that the money is well used. But in the same way, reforms are pointless if budgets are so restricted that little serious research can be performed — as is now the case in Bulgaria.

The Bulgarian government, together with its scientists, must urgently create a long-term scientific plan for the country, and a strategy to put the plan into operation and ensure that it is successful. It cannot afford to reject the European Union philosophy of a future centred on a knowledge-based economy.

In the meantime, it needs to restore budgets for science and universities to levels that allow them to function properly, and delay plans to break up the academy. Only when a proper long-term strategy is in place will the government know what it needs to do about its research institutes and their budgets. If it needs inspiration in this, it need only look north to its neighbour Romania. The contrast between the age of wisdom and the age of foolishness is clear. ■

First do no harm

Simple tools to diagnose mental illness should not be offered without sound supporting evidence.

Incorrect diagnosis of people with psychiatric disorders has far-reaching implications. Miss the manic phases of people with bipolar disorder, for instance, and, rather than the mood-stabilizers they need, they might be given antidepressants. The drugs could make them 'hypomanic' — a state in which they might spend money recklessly, invest irrationally and jabber incessantly so that friends and employers no longer want them around — and tip them into even more extreme bipolar cycles. Meanwhile, if the hallucinations of a person with schizophrenia don't become apparent during analysis, the patient may likewise be diagnosed as depressed, be given antidepressants and go on to become even more withdrawn. Unfortunately, misdiagnosis happens all too often — in around 70% of cases of bipolar disorder, according to some estimates. And such mistakes often go uncorrected for years.

The problem is that, in the absence of clear-cut biological markers for such disorders, doctors depend on subjective examinations to fit patients into poorly defined categories. The uncertainty inherent in these diagnoses is a menace, and not only for patients. The problem affects pharmaceutical companies working to develop new drugs, insurance companies trying to determine coverage, health-care systems and employers. Researchers are desperate for objective diagnostic markers to replace subjective examinations. In their search they have studied genetics and investigated tools such as electroencephalography and functional magnetic resonance imaging. Despite many claims and limited successes, so far none of these findings has been replicated consistently enough to merit widespread clinical use.

Last year, doctors in Japan started using a relatively new technique — near-infrared spectroscopy (NIRS) — to distinguish between schizophrenia, depression, bipolar disorder and normal mental-health states (see page 148). It is not difficult to see why this was a popular

move. Doctors like it because it is easy to use. Patients like its objective nature, and that it provides them with physiological evidence of a disorder. And officials at the health ministry are happy because it represents a public success of their drive to innovate.

In Japan, NIRS diagnosis is one of dozens of advanced medical techniques offered to patients — at their own expense — despite not having gone through the clinical trials needed for approval by national health insurance. But is it ready for the clinic? Most scientists contacted by *Nature* do not think so. The patient groups of the supporting studies were small. The tests have not been reproduced in various clinical settings as one might hope. There is no international consensus on how best to measure NIRS parameters, much less a clear consensus on how to apply them to mental health. And if it is not ready, the same advantage that has patients lining up for it — the authority of an objective, physiological measure, the air of the incontrovertible — will become an obstacle. A misdiagnosis that carries the authority of an 'objective' measurement will probably be even more difficult to overturn.

The doctors who offer the technique say that it is only an aid, just one tool in their toolbox. Indeed, they are following Japan's advanced medical technology protocol properly and offering, in good faith, a diagnostic test that they believe works. Their attempts to use it seem sincere, and not motivated by profit. Still, if the technique has not been clinically validated, patients should not be paying for it, particularly given the challenging scientific landscape of mental-health problems. Japan's advanced medical-technologies programme is blurring the line between protocols that have been properly validated and those that have not. More rigorous verification methods are needed, starting with multiple blind trials in large patient groups.

NIRS does show promise. It is easy, quick and, perhaps combined with other diagnostic techniques, could be a powerful tool, if the right validation studies are done. If the clinics that offer the technique are so sure that it works, then they should present the supporting evidence to prove it.

Meanwhile, other neuroscientists should continue the search for more conventional biomarkers. To offer better care to people with mental-health problems is a noble motive, but one that needs to have science alongside desire at its heart. ■

➤ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunq



University cuts show science is far from saved

Scientific leaders have been too quick to praise the reprieve for research money, says **Colin Macilwain**. The slashing of teaching funds will do real damage.

In countries where the economic crisis has hit hardest, science has not done badly — so far. But universities from Bologna to Berkeley face an almost existential crisis. While governments defend research spending, they are simultaneously slashing public funding for universities, where most research takes place.

The reaction from science lobby groups and figureheads in the scientific community to this situation has been bafflingly cheerful. Either they have lost touch with what's happening on the ground, or else they are preoccupied with flattering politicians for 'saving science' — when politicians are actually cutting the very ground from underneath it. Most researchers know what is really going on, however, because they work in the universities where overall budgets are under the hammer.

Science today is so thoroughly embedded in universities that the line between the two has become difficult to discern. And research in universities requires solid undergraduate and graduate learning and teaching. It is foolhardy to weaken this foundation, because the modern research university is built on the energy and ideas of students. Students are not customers of a university; they are its very soul. The idea that research will prosper while teaching and learning decay is a dangerous fallacy.

The failure of many in the science establishment to pursue this point is most visible in Britain, where money for research and teaching comes from the same pot: the Department for Business, Innovation and Skills. In the autumn spending review, warmly praised by many who claim to speak for UK science, this department saw its budget cut more steeply than any other big-spending arm of the UK government — by 8% a year for four years.

When the cuts were announced, John Beddington, the government's chief scientific adviser, joined other officials in boasting that science had been protected, after Treasury officials were persuaded of its worth (see *Nature* 467, 1017; 2010). But the Treasury hadn't given an inch. Science was protected purely by eviscerating public support for university teaching in England.

The reaction of Wellcome Trust head Mark Walport was typical. "I am delighted that the government has recognised the huge importance of science," he said. "The government has listened to the voices of the science community who argued that continued investment in science was vital to the United Kingdom's future success. It is now up to the science community to ensure it delivers on this crucial vote of confidence."

One problem with this promise is that it isn't within the power of the universities, or scientists, to deliver a competitive economy.

As Geoffrey Boulton of the University of Edinburgh and Colin Lucas, former vice-chancellor of the University of Oxford, have pointed out,

governments have started to make crazy assumptions about the ability of universities to deliver innovative companies and successful economies. In a 2008 League of European Research Universities paper, *What are Universities For?*, the duo argued that the thrust of higher-education policy in many countries is "squeezing out diversity of function and undermining teaching and learning". Among policy-makers, they warned, "slipshod thinking about universities is leading to demands that they cannot satisfy, while obscuring their most important contributions to society and undermining their potential".

Boulton and Lucas were talking mainly about Europe, but there are related problems in the United States. University management there is too often obsessed with building grandiose labs, to be financed by over-heads on future research grants they expect to win from the National Institutes of Health (NIH) (B. Alberts *Science* 329, 1257; 2010). With major expansion at the NIH over, and state government support for

teaching in rapid decline, many institutions are now locked in a futile battle to fill these white elephants, creating what biochemist Kenneth Mann of the University of Vermont in Burlington has dubbed "a toxic, uncertain environment" for students.

With the long-term decline of top-class independent or corporate laboratories, almost all Nobel prize-level science is now done at universities. And the greatest universities, starting at the top with Harvard, increasingly define themselves chiefly in terms of their scientific prowess — or, more prosaically, by the amount of research funding they can attract.

When the universities were doing well — and in many parts of the world, they have just enjoyed decades of expansion — the concentration of scientific research within their walls was more or less entirely beneficial. When the economic storm struck in 2008, the ride came to an abrupt end. Now,

as Western governments attempt to maintain investment in science as a route to innovation and industrial development, they are undermining support for students and the quality of their education. Instead of joining with students and teaching staff elsewhere in academia in protest, too many scientific leaders have stood aloof. (Martin Rees, until this month the president of the Royal Society in London, is a notable exception.) Strategically, this approach is a disaster in waiting.

China and India know this and are building universities from the ground up, with a firm emphasis on student education as their bedrock of energy and ideas. In the United Kingdom and elsewhere, these foundations are being demolished, and students drowned in debt, to keep researchers' grants flowing. It can only end badly, and more in the scientific establishment should have the courage to say so. ■

Colin Macilwain is a contributing correspondent with *Nature*.
e-mail: cfmworldview@gmail.com

THE IDEA THAT
RESEARCH
WILL PROSPER
WHILE TEACHING
DECAYS
IS A
DANGEROUS
FALLACY.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/wvqplv

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

MICROBIOLOGY

'Jet lag' weakens malaria parasite

The periodicity of the life cycle of malaria parasites during host infection suggests that these organisms are regulated by circadian rhythms. Aidan O'Donnell at the University of Edinburgh, UK, and his team found that parasites whose rhythms were not synchronized with those of their hosts were less able to replicate in, and transmit out of, the hosts.

The researchers 'set' the circadian rhythms of mice and the rodent parasite *Plasmodium chabaudi* by keeping them in one of two rooms with opposing 12-hour light-dark cycles. They then infected the mice with parasites that were either in or out of sync with the animals' own cycles.

When mouse and malaria rhythms were the same, parasite densities during the replication and transmission stages of infection were double those seen for parasites that were out of sync with the mice. **Proc. R. Soc. B** doi:10.1098/rspb.2010.2457 (2011)

ECOLOGY

The effects of opossum shrimp

An analysis of more than a century's worth of ecosystem data has revealed how the

introduction of a single species of shrimp to a US lake led to a cascade of changes in the food web.

Bonnie Ellis at the University of Montana's Flathead Lake Biological Station in Polson and her colleagues examined published data from Flathead Lake (pictured) dating back to the end of the nineteenth century. They found that the opossum shrimp (*Mysis diluviana*), which was introduced into the lake in the mid-1980s, became a food source for the previously

introduced but theretofore unobtrusive lake trout, which now dominates the lake.

The lake trout went on to eat all of the kokanee salmon, depriving eagles of their annual spawning kokanee feast. In addition, the shrimp consume large zooplankton, so small zooplankton now dominate. Because the latter do not consume as much algae, algal levels have increased, leading to a 21% rise in photosynthesis. **Proc. Natl Acad. Sci. USA** doi:10.1073/pnas.1013006108 (2011)

CELL BIOLOGY

Best of two microscopes

Electron microscopes allow cell biologists to visualize the tiniest of cellular features, but struggle to locate rare features or events. Fluorescence light microscopy (FM) is well suited to this task, but its resolution is low. So Marko Kaksonen, John Briggs and their colleagues at the European Molecular Biology Laboratory in Heidelberg, Germany, combined the two modalities to image rare



CLIMATE CHANGE ECOLOGY

Butterflies break out earlier

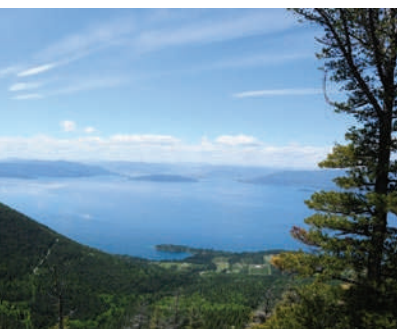
As Earth's climate changes, many butterfly species are emerging — from cocoons or from hiding — earlier in spring. Researchers have identified traits in UK butterflies that predict the largest shifts seen in emergence times over the past 30 years.

Sarah Diamond, then at the University of North Carolina, Chapel Hill, and her colleagues analysed a data set on UK butterflies such as the speckled wood (*Pararge aegeria*, pictured) during a period in which the country's spring

temperatures rose by 1.5 °C. Butterflies that eat a lower diversity of plant species as caterpillars showed larger shifts in emergence time. The authors suggest that the butterflies may be tracking changes in their host plants' annual schedules.

Bigger shifts were also apparent in butterflies that overwinter as adults rather than as pupae or larvae. These species may respond more quickly to warm temperatures, the authors say. **Ecology** doi:10.1890/10-1594.1 (2011)

J. STANFORD



R. COOPER/NATURE PICTURE LIBRARY

features with a precision of less than 100 nanometres.

Using yeast cells, the authors show that most signals from fluorescently tagged proteins are still detectable even after the cells have been embedded in resin and sectioned for electron microscopy (EM). To correlate signals from FM and EM, the researchers introduced small microspheres visible in both modalities into the samples. From these, they were able to map features found in FM to images obtained with EM.

The team used the technique to image single HIV particles on mammalian cell surfaces.

J. Cell Biol. doi:10.1083/jcb.201009037 (2011)

CANCER GENETICS

One catastrophe, many mutations

Cancer is thought to develop through the gradual accumulation of mutations across the genome. But a small percentage of cancers may instead come about through a one-off onslaught of tens to hundreds of rearrangements to a single chromosome.

Peter Campbell at the Wellcome Trust Sanger Institute in Cambridge, UK, and his colleagues analysed the genomes of ten patients with leukaemia. One patient had several distinct patterns of genomic rearrangement, such as a concentration of mutations in one section of one chromosome, and gene copy numbers — often raised in cancer — that fluctuated between just one and two. Further analysis of more than 3,000 other patient samples revealed similar patterns in 2–3% of people with cancer.

From modelling work, the authors conclude that such patterns are consistent with a chromosome shattering into tens to hundreds of pieces, many of which are then pieced together by the cell's DNA-repair machinery, leaving mistakes. Such a catastrophe could lead to many cancer-causing mutations.

Cell 144, 27–40 (2011)

MATERIALS SCIENCE

Spinning yarns of nanotubes

Materials such as metals and semiconductors have been spun into yarns that can be woven, sewn and tied in knots, raising the possibility of new types of textiles with electronic capabilities.

Ray Baughman and his co-workers at the University of Texas at Dallas used a low concentration of multi-walled carbon nanotubes as the 'host' material to provide durable, flexible and weavable yarns. A high concentration of 'guest' material added the desired functionality. The researchers fabricated 100-micrometre-wide yarns by twist-spinning specific combinations of the two components, a method that overcomes limitations, such as lack of material durability and functionality, of previous approaches.

Using lithium iron phosphate (LiFePO₄) as the cathode (pictured) and graphene nanoribbon yarn as the anode, the authors describe how to create a weavable yarn with promising lithium-ion battery functionality.

Science 331, 51–55 (2011)



NEUROIMMUNOLOGY

Immune input for retinal repair

Healing of the eye's retina after injury is aided by immune cells that dampen local inflammation and boost the retina's capacity for repair.

Michal Schwartz at the Weizmann Institute of Science in Rehovot, Israel, and her team monitored the influx of cells called macrophages from the blood into the damaged retinas of mice. They found that the immune cells flooded

COMMUNITY CHOICE

The most viewed papers in science

MATERIALS SCIENCE

Graphene layers made to order

HIGHLY READ
on *apl.aip.org* in
November 2010

By flinging carbon atoms at a nickel film, researchers have succeeded in generating specific numbers of layers of graphene

— atomically thick sheets of carbon that promise to surpass silicon as a material in electronic devices. Different thicknesses have markedly different properties, but controlling thickness is a challenge for current production techniques, with some able to generate only single layers.

Jene Golovchenko and his colleagues at Harvard University in Cambridge, Massachusetts, accelerated carbon atoms in an electric field and deposited doses corresponding to the amount needed for one-, two- or three-layered graphene onto nickel sheets. When heated, the carbon formed graphene of the specified thickness.

Because the underlying technology used in this method is already widespread in the fabrication of semiconductors, it could be easily used for large-scale graphene production.

Appl. Phys. Lett. 97, 183103 (2010)

the retinas, with most homing in on the area of damage — the layer containing the retinal ganglion cells (RGCs). Boosting macrophage levels enhanced the survival of RGCs and triggered proliferation of retinal progenitor cells. The macrophages also altered the levels of immune regulators to decrease inflammation in the damaged retina.

The authors did not find any signs of newly generated mature retinal neurons. They therefore suggest that the progenitor cells provide a protective environment, for example by modulating the immune response, rather than replacing damaged cells.

J. Exp. Med. doi:10.1084/jem.20101202 (2011)

MOLECULAR EVOLUTION

The birth of a gene

How new genes emerge while others decay could be explained, in part, by an unusual means of gene duplication.

Ichizo Kobayashi at the University of Tokyo and his team compared the

complete genome sequences of ten strains of the bacterium *Helicobacter pylori*, which causes stomach ulcers, from people living around the world. From this, they reconstructed the evolutionary history of the organism's chromosome organization.

Some strains contain a single copy of a host-interaction gene, whereas others have two. The duplicate gene and the DNA between the two have been flipped around. The researchers think that a single process copied the gene while inverting the DNA. In some strains, this mechanism seems to have interrupted a host-interaction gene, causing it to decay, whereas in other strains a new gene has evolved as a result of the process.

The same mechanism could occur in cancer cells, creating abnormal numbers and arrangements of genes, the researchers say.

Proc. Natl Acad. Sci. USA
doi:10.1073/pnas.1012579108 (2011)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

UK health research

The UK government has welcomed a call for simpler medical-research regulations. An 11 January report from the Academy of Medical Sciences in London recommends the creation of a new health-research agency to offer a single body for governance and ethics checks of new clinical trials. It also calls for reform of the European Union's much-criticized Clinical Trials Directive legislation, and for improvements in the UK National Health Service's attitude towards research. Health secretary Andrew Lansley said in a statement that the government would "consider carefully how to implement" the academy's advice. See go.nature.com/vi8rcp for more.

Data-sharing pact

An international coalition of 17 public and private funders has endorsed a statement calling for greater sharing of public-health research data.

SOUND BITE

“Is it possible that he was wrong, but not dishonest ... ? No.”

The British Medical Journal (BMJ) on Andrew Wakefield, who is alleged, in a BMJ investigation published last week, to have manipulated patient medical histories in a 1998 paper in *The Lancet* linking childhood vaccines to autism and bowel disease.

Source: *Br. Med. J.* **342**, c7452 (2011).



T. BLACKWOOD/AFP/GETTY IMAGES

Australian floods pose threat to reefs

A mix of pesticides, fertilizers and other flood run-off could threaten the biodiversity of Australia's Great Barrier Reef. Heavy rains since late December have resulted in flooding in the state of Queensland, and satellite imagery

shows large quantities of sediment streaming out of the Burdekin River and towards southern stretches of the reef, resulting in algal blooms. Previous floods have also caused population spikes in coral-eating starfish.

Signatories include the US National Institutes of Health, Britain's Wellcome Trust and the Bill & Melinda Gates Foundation. The statement calls on funders to establish "equitable", "ethical" and "efficient" data-sharing policies.

Food costs soar

World food prices reached a record high last month, according to a report released on 5 January by the Food and Agriculture Organization of the United Nations in Rome. The Food Price Index, which covers dairy products, meat, sugar, cereals and oilseeds, averaged 214.7 points for December 2010, up 4.2% from 206 points a month earlier and slightly above the previous peak of 213.5 points in June 2008, when food prices triggered riots in a number of countries.

This is the highest the index has reached since it was created in 1990. See go.nature.com/tf17dz for more.

University law

Hugo Chávez, the Venezuelan president, has vetoed a law that would have placed university budgets and elections in the hands of 'communal councils' of local citizens. The approval of the University Law late last year by the country's outgoing National Assembly sparked protests by students and researchers. The law is now with the newly elected assembly, whose members took their seats on 5 January.

Converging science

The integration of biology with physical sciences and engineering is a research model that deserves to be encouraged by specific

facilities and more funding, a group of high-profile scientists at the Massachusetts Institute of Technology in Cambridge have urged. They published a report on 4 January that dubbed such cross-disciplinary work 'convergence', and argued that funding agencies should reorganize themselves to better support the concept. See go.nature.com/o7xus2 for more.

Oil-spill report

The Gulf of Mexico oil spill was caused by a string of avoidable mistakes, says the US presidential panel tasked with investigating the blowout on 20 April 2010 and the resulting spill. The explosion on the Deepwater Horizon rig killed 11 workers and led to the release of about 5 million barrels of oil into the Gulf. The

S. WALSH/AP

national oil-spill commission, whose full report is out this week, blamed the spill on errors by the companies involved, including BP — who owned the well — and contractors Transocean and Halliburton, as well as lax government regulators.

PEOPLE

FDA official leaves

The deputy commissioner of the US Food and Drug Administration (FDA) leaves his post this week to head the state of Maryland's health department. Joshua Sharfstein arrived with the administration of President Barack Obama in 2009 to help steer an agency that critics said had become too friendly with the drug, device and food manufacturers it regulates. During Sharfstein's tenure, the FDA curbed prescription of the diabetes drug Avandia, admitted it erred in approving a device used in knee surgeries and effectively banned the sale of drinks that are manufactured to contain both alcohol and caffeine.

Representative shot

A lone gunman's assassination attempt on congresswoman Gabrielle Giffords (Democrat, Arizona), who chaired a subcommittee that oversees NASA, is sending ripples through the science and



space communities. Giffords (pictured) was shot in the head at a public meeting on 8 January, in an attack that killed 6 and wounded 14 other people. As *Nature* went to press she was in critical condition. She was elected to the House of Representatives in 2007 and is also an active supporter of renewable energies such as solar power and of the America COMPETES Act, which encourages increased science funding.

BUSINESS

Stem-cell trial

The US Food and Drug Administration has given the green light for a third human trial using human embryonic stem cells. Advanced Cell Technology (ACT), which last November won approval to use the stem cells to treat a rare form of juvenile blindness called Stargardt's disease,

announced on 3 January that it had received the go-ahead to use the cells in a trial of adults with a related but far more common condition: age-related macular degeneration. The company, based in Santa Monica, California, said it would also pursue approval for a trial in Europe. See go.nature.com/xyjyty for more.

Rare-earth rules

China is preparing new standards to govern the production of rare-earth metals, according to state media reports. The metals — scandium, yttrium and the lanthanides — are key to electronic equipment and green technology, in part because of their use in batteries. China controls the lion's share of world production and last year hinted that it could raise taxes and reduce exports. The regulations could be officially released as soon as February. See go.nature.com/lqnt3d for more.

Journal launched

A new open-access journal, *Scientific Reports*, will release its first batch of articles in June 2011. The peer-reviewed online publication, announced by Nature Publishing Group on 6 January, will accept biology, chemistry, physics and Earth sciences papers deemed scientifically sound

COMING UP

19–21 JANUARY

Environmental scientists and policy-makers meet in Washington DC to discuss how to protect and restore the world's oceans, at a conference convened by the US National Council for Science and the Environment.

go.nature.com/6w465b

and original without regard to the significance of the research, a policy similar to that of the journal *PLoS ONE*. *Scientific Reports* will charge researchers for publication: US\$1,350 for manuscripts accepted in 2011 and \$1,700 for those accepted in 2012.

RESEARCH

Tevatron to close

The US Department of Energy has decided not to fund a three-year extension of the Tevatron, the particle collider at the Fermi National Accelerator Laboratory in Batavia, Illinois. It will close at the end of the 2011 fiscal year. See page 141 for more.

Planck data released

Scientists with the Planck space mission have unveiled its first results. Launched in 2009, the €600-million (US\$770-million) spacecraft is designed to study the cosmic microwave background, the radiation released just after the Big Bang. The satellite has also discovered scores of objects and evidence of primordial galaxies that have star formation rates 10–1,000 times that of the Milky Way today, scientists said at a press conference on 11 January. Planck's data collection continues, with a complete microwave map expected in January 2013.

NATURE.COM

For daily news updates see:

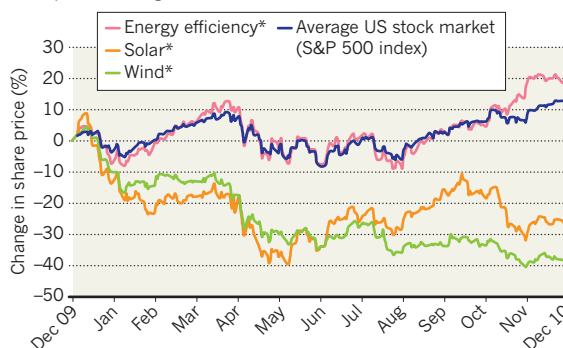
www.nature.com/news

BUSINESS WATCH

Clean energy stocks tracked by the WilderHill New Energy Global Innovation Index fell 14.6% overall in 2010, mainly because of the poor performance of the index's solar and wind companies (see chart). Ethan Zindler of analysts Bloomberg New Energy Finance says a race to drive down the price of solar panels and wind turbines was largely responsible — good for the market, but bad for sales figures. Investor uncertainty over subsidies also contributed to the decline.

CLEAN-ENERGY SLUMP

Shares of solar and wind companies dipped even though US stocks rose overall in 2010. But companies working on energy efficiency — from power management to electric vehicles — fared better.



*Collation of companies involved in these sectors, as tracked by the WilderHill New Energy Global Innovation Index

SOURCE: BLOOMBERG NEW ENERGY FINANCE/WILDERHILL NEW ENERGY INDEX

NEWS IN FOCUS

PHYSICS Accelerator closing dooms US bid for Higgs particle **p.141**



ASTRONOMY Planet-hunting telescope finds its first rocky world **p.143**

NEUROSCIENCE Can brain imaging diagnose mental health? **p.148**

ECOLOGY Yellowstone park braces itself for a warmer future **p.150**

C. KEANE/REUTERS



Duke University allowed a controversial set of clinical trials to continue despite serious concerns about the validity of the data on which they were based.

ETHICS

Cancer trial errors revealed

University officials admit data withheld from review panel before misconduct charges arose.

BY EUGENIE SAMUEL REICH

It was a weekend that Michael Cuffe, vice-president for medical affairs at Duke University in Durham, North Carolina, says he will never forget. It began on 16 July 2010, when Cuffe learned of a damning revelation in *The Cancer Letter*, a Washington DC-based publication with a reputation for probing controversial topics in cancer research. A story in that day's issue alleged that Anil Potti, a cancer geneticist whose data had been used to design three clinical trials then under way at Duke, had lied on multiple federal grant applications, falsely claiming, among other things, to have been a Rhodes scholar.

That was bad news, but for Cuffe and for Sally Kornbluth, Duke's vice-president for research, that was not the worst of it. Months earlier, the same trials had been suspended after critics raised questions about the underlying science.

Yet Cuffe and Kornbluth had decided to restart them when a review panel seemed to validate Potti's method. The allegations that Potti, who worked at Duke's Institute for Genome Sciences and Policy (ISGP), had padded his CV changed everything. "When it comes to light that someone may have been less than honest in one aspect of their professional life, one begins to wonder whether they have been less than honest in another aspect," says Kornbluth. That weekend, she and Cuffe suspended the trials once again, and initiated a misconduct investigation that is still ongoing. Potti, who could not be reached for comment, resigned from the ISGP in November and took full responsibility for irregularities in his data.

Now, in response to information obtained by *Nature* under the US

Freedom of Information Act, Kornbluth and Cuffe have offered their account of the mistakes that led the trials to be restarted even after they learned of potential flaws in the underlying data. The affair will have an impact beyond Duke, as the Institute of Medicine, part of the US National Academies in Washington DC, begins to examine research on genome-based patient testing. Originally commissioned to investigate Duke's controversial trials, the institute's US\$687,000 study is now expected to focus on providing broader recommendations for the design of clinical trials that similarly use genomic data from individual patients to tailor therapy.

ACADEMIC DISAGREEMENT

Starting in 2006, Potti and his colleagues at Duke had filed patent applications and published papers^{1,2} describing genomic predictors — computer algorithms that take gene-expression data from a cancer cell and predict whether ▶

► **NATURE.COM**

To read the external review panel's report, visit: go.nature.com/umtymu

► the cancer will be sensitive to a particular therapy (see 'How events unfolded').

Duke began three clinical trials based on the technology, ultimately enrolling around 110 patients with lung or breast cancer. But Potti's work was under fire. Keith Baggerly and Kevin Coombes, biostatisticians at the University of Texas M. D. Anderson Cancer Center in Houston, published a technical comment³ and a paper⁴ saying that they had been unable to replicate Potti's findings. As a result, officials at the National Cancer Institute (NCI) in Bethesda, Maryland — which had received a proposal from Duke to start a fourth clinical trial based on Potti's work — contacted the university to request a review. The trials were suspended in September 2009 and a panel of anonymous reviewers, external to Duke, began its work.

While the review was under way, Baggerly and Coombes analysed data that Potti (and his co-authors, none of whom have been implicated in any misconduct) had posted online. The biostatisticians wrote to Kornbluth and Cuffe, pointing out that the data did not match the raw data in the public database from which they were supposedly sourced. "They had numbers with labels that the Duke group said applied, but the labels were wrong," says Baggerly. Yet in December 2009, Duke's review panel concluded that Potti's work was valid, and the trials were restarted.

In an NCI report obtained by *Nature*, Duke's external reviewers say that they can replicate the results using data provided by Potti, but seem unaware of any doubts about the data. Kornbluth and Cuffe admit that, in consultation with John Harrelson, who was acting as chairman of Duke's Institutional Review Board, they decided not to forward the latest communication from Baggerly and Coombes to the rest of the board or the external reviewers.

Kornbluth denies trying to protect Potti. "Our motivation was not to protect [him], it was to give [him] complete fairness," she says.

They judged that the panel could be improperly biased if it received Baggerly and Coombes's letter, says Cuffe. He and Kornbluth emphasize that, in the early stages of the case, before misconduct allegations were made, they believed that Potti was providing them with "full and truthful answers". But Cuffe acknowledges that the possibility of misconduct may have crossed their minds. "It's not like we had never considered this," he says. If the situation arose again, he says that he would forward "every shred" of evidence to a review panel.

Baggerly regrets that the document wasn't forwarded. "A lot of the back and forth of the past year could have been avoided," he says.

INCREASED SCRUTINY

Even before Potti's CV padding came to light, doubts about his work were re-emerging. Last April, officials at the NCI determined that they were part-funding the clinical trials through a grant to Potti. NCI documents show that officials weren't reassured by the review panel's report, having found themselves unable to replicate Potti's results using data from the public database. The institute asked Duke to provide it with the information needed for replication, but that process was overtaken by the allegations from *The Cancer Letter*.

Kornbluth and Cuffe say that Duke is doing its best to learn from the case, and has set up a committee to decide what checks should be in place on basic research before starting clinical trials. Meanwhile, the Institute of Medicine has begun its study, which is expected to last 15 months, and will cover a variety of 'omics-based' tests, including genomics, proteomics and metabolomics.

David Brooks, chief medical officer of Generation Health, a pharmacogenetics testing company in Upper Saddle River, New Jersey, estimates that more than half of cancer clinical trials involve some kind of genomic testing, so the institute's review may have widespread relevance. Richard Simon, a biostatistician at the

NCI, says that one consideration in planning trials is whether genomics data should be used to allocate patients to one therapy or another. An alternative study design is to allocate therapy randomly and have the genomics studied at the same time to see how the two correlate. "It avoids the ethical issues," says Simon.

Back at Duke, other researchers are under scrutiny. Steven McKinney, a statistician at the British Columbia Cancer Research Centre in Vancouver, Canada, has written to the Institute of Medicine to raise the concern, not about misconduct, but that statistical methods used in other Duke studies overlap with a method used by Potti. McKinney told *Nature* that he is troubled by one study co-authored by Geoffrey Ginsburg, a physician at the ISGP. In the study⁵, approved by Duke's Institutional Review Board and review boards at four further institutions, 57 people were infected with cold or influenza viruses. Gene-expression data were then obtained to see whether it was possible to retrospectively 'predict' which virus they had been infected with — information that could aid the prompt detection of a pandemic. McKinney is concerned that the statistical method being used does not seem to have been proven.

Ginsburg says that "in contrast to the concerns around the cancer work we have gone to great lengths to replicate the results using a variety of statistical methodologies". He and his co-authors will look in detail at McKinney's concerns. Kornbluth echoes this sentiment. "This is something we're going to take seriously and look at very closely," she says. ■

1. Potti, A. et al. *Nature Med.* **12**, 1294–1300 (2006).
2. Hsu, D. S. et al. *J. Clin. Oncol.* **25**, 4350–4357 (2007).
3. Coombes, K. R., Wang, J. & Baggerly, K. A. *Nature Med.* **13**, 1276–1277 (2007).
4. Baggerly, K. A. & Coombes, K. R. *Ann. Appl. Stat.* **3**, 1309–1334 (2009).
5. Zaas, A. K. et al. *Cell Host Microbe* **6**, 207–217 (2009).

HOW EVENTS UNFOLDED

2006

Anil Potti (pictured), a cancer geneticist at Duke University in Durham, North Carolina, and others file patent applications on the idea of using gene-expression data to predict sensitivity to cancer drugs. Potti is first author on a paper in *Nature Medicine*¹.



2007

Potti is last author on a paper in the *Journal of Clinical Oncology* (JCO)². Duke begins three clinical trials to test Potti's predictors in patients with breast or lung cancer.

SEPTEMBER 2009

Keith Baggerly (pictured) and Kevin Coombes, statisticians at the University of Texas M. D. Anderson Cancer Centre in Houston, publish a paper in *Annals of Applied Statistics*³ stating that they could not replicate Potti's claims. Duke suspends the trials and asks a review panel to investigate.



NOVEMBER 2009

Potti places data underlying the JCO paper online. Baggerly writes to Sally Kornbluth (pictured), Duke vice-dean for research, and Michael Cuffe, Duke vice-president for medical affairs, to point out differences from raw data.

DECEMBER 2009

An unredacted copy of the report by Duke's review panel, later obtained by *Nature*, shows that the panel replicated Potti's claims using his data, but were unaware that those data contained discrepancies.



JANUARY 2010

Duke restarts clinical trials.

JUNE 2010

Harold Varmus (pictured), director of the National Cancer Institute in Bethesda, Maryland, asks the Institute of Medicine to review Duke's trials.

JULY 2010

The Cancer Letter reveals that Potti made false claims about his CV. Trials are suspended and an investigation begins.



NOVEMBER 2010

JCO paper is retracted. Duke closes the trials permanently. Potti resigns.

DECEMBER 2010

Institute of Medicine study begins, but will now focus more generally on criteria for genomics predictor.

JANUARY 2011

Nature Medicine paper is retracted.

HIGH-ENERGY PHYSICS

Tevatron faces final curtain

Particle accelerator to be switched off this year, as lack of funds spells the end for US bid to capture Higgs particle.

BY EUGENIE SAMUEL REICH

Depending on who you talk to, it is either a disappointing blow or a clean break heralding an exciting new era. After much debate, officials at the US Department of Energy's Office of Science revealed this week that they have decided not to extend funding for the Tevatron, the proton-antiproton collider at Fermilab in Batavia, Illinois, by an additional three years. The decision means that the first glimpse of the long-predicted Higgs particle, thought to endow other particles with mass, will probably be achieved by the Large Hadron Collider (LHC) at CERN, Europe's particle-physics lab near Geneva in Switzerland.

The decision was explained in a letter sent on 10 January by Bill Brinkman, director of the Office of Science, to Melvyn Shochet, a physicist at the University of Chicago, Illinois, and chairman of the energy department's High Energy Particle Advisory Panel (HEPAP). In October 2010, with the LHC suffering from delays, HEPAP had recommended that the US machine be extended beyond its planned 2011 closure if extra funding of US\$35 million could be found. It couldn't, says Brinkman. "Unfortunately, the current budgetary climate is very challenging and additional funding has not been identified," he writes in his letter. He adds that the Tevatron will shut down this year as planned.

The decision is a blow for the 1,200 physicists who work on experiments at the Tevatron, the world's second most powerful accelerator after the LHC, as well as for particle-physics theorists at Fermilab and elsewhere who would have enjoyed working with the data. "I feel disappointed," says Pier Oddone, Fermilab's director. "We all would have wanted to see the Tevatron continue."

Rick Van Kooten of Indiana University Bloomington, who recommended the extension to the Tevatron in his role as chairman of Fermilab's Physics Advisory Committee, says he's also disappointed at the decision, but is convinced that sincere efforts were made to find the funding.

Leaders of Tevatron experiments made the case for extension last year, arguing that the machine's smooth running, combined with



The Tevatron's 6.3-kilometre-long ring will stop colliding particles this year rather than in 2014.

technical problems at the LHC, provided the Tevatron with a shot at discovering the Higgs, and the glory that would go with it. The two advisory panels also agreed that the physics case was compelling. But Charles Baltay, a physicist at Yale University in New Haven, Connecticut, and the chairman of P5, the HEPAP subpanel that considered the request, says this week's decision not to go ahead is consistent with the P5 recommendation that the Tevatron extension be funded only by adding to the high-energy physics budget, and not by taking funds away from other US particle-physics experiments. "I'm glad they made a clean, unambiguous decision," he says. "One concern was a drawn-out process."

In a way, the decision can be seen as a vote of support for a programme mapped out by P5 in 2007 and 2008. According to that programme, the United States should concede the realm known as the 'energy frontier', involving the highest-energy particle collisions, to the LHC, instead focusing its domestic efforts on 'the intensity frontier', achieving the greatest number of particle collisions per second. The latter is ideal for studying rare processes, and so P5 recommended investment in a slew of experiments to do just that. They include Fermilab's Mu2e, which will study the rare decay

of muons to electrons, and NOvA and the Long Baseline Neutrino Experiment, which will pin down the mass and other properties of elusive neutrinos, expected to have relevance for explaining the asymmetry between matter and antimatter in the Universe.

Mark Messier of Indiana University Bloomington, co-spokesman for NOvA, says the decision makes things easier for experiments that would have competed for resources had the extension been approved. "We're back on the original course. There's something that we now don't have to deal with," he says. The Tevatron extension would have delayed the start of NOvA, because the neutrino experiment is designed to use the Tevatron's cast-off recycler, a particle-storage ring.

Robert Roser, co-spokesman for one Tevatron experiment, the Collider Detector at Fermilab, says that physicists on the experiment will now focus on ensuring the most successful end to the current run. "We raised the bar for the LHC. We're very proud of our accomplishments." The analysis of data from the machine is expected to continue for the next two to three years.

Oddone says that the exact switch-off date will become clear once the US Congress passes the fiscal year 2011 budget. ■

"I'm glad they made a clean, unambiguous decision. One concern was a drawn-out process."

Science fortunes of Balkan neighbours diverge

Romanian researchers' prospects improve, whereas scientists in Bulgaria face a bleak future.

BY ALISON ABBOTT

When Bulgaria and Romania joined the European Union (EU) in 2007, ambitious scientists there hoped that it would swiftly disperse the lingering legacies of 40 years of communist rule. Science had been blighted by a lack of competition, minimal funding and an old guard so desperate to cling on to its privileges that fresh blood was denied the opportunity to build a research career.

Four years later, these neighbouring cash-strapped countries on the southeastern edge of the EU remain close neighbours at the bottom of EU league tables of research expenditure and research output statistics (see 'The spending gap'). But their scientific fates may be set to diverge dramatically.

Last year, Bulgaria slashed the budget of its Academy of Sciences by 38% to 75 million leva (US\$50 million), forcing some scientists to take unpaid leave. It reportedly reduced the budget of its granting agency by more than half, and will maintain those funding levels in 2011. It is also poised to make organizational changes that scientists fear will foster mediocrity.

Over the same period, Romania has trimmed funding of its Academy of Sciences by just 3%, and increased the budget of its granting agency by 44% to 1.5 billion lei (US\$0.5 billion). In a drive to improve the standard of Romanian science, the government last week approved a hard-won law that

enforces strict quality control on universities and introduces tough rules for funding evaluation and peer review.

Romanian scientists working outside the country say that the changes give them hope of some day being able to continue their research careers back home. Meanwhile, the Bulgarian diaspora despairs.

These opposing outlooks show how joining the EU does



Students at Sofia University have joined protests over Bulgaria's plans to cut research and education funds.

not guarantee a successful scientific transformation, despite the access it offers to outside money and expertise, and the EU's focus on innovation. For science to flourish, national political will is crucial.

A TALE OF TWO COUNTRIES

After Bulgaria and Romania became democracies in 1990, both attempted to reform their research organizations and universities along Western lines. Early on in the process, they gave autonomy to their respective scientific academies, learned societies that also run extensive networks of research institutes and that were under state control in Soviet times. But little reform accompanied that autonomy, a matter of political concern for many years.

In November 2010, the Bulgarian science and education minister, Sergei Ignatov, proposed an amendment to the law governing the Bulgarian Academy of Sciences (BAS) that would bring its institutes back under the control of the ministry, with each institute funded separately, leaving the academy as a learned society. The amendment, expected

to be approved in the next month or so, does not outline how the ministry would safeguard the scientific objectives of the institutes. Scientists fear that the institutes will be closed or deprived of the opportunity to do internationally relevant research.

Enraged BAS scientists were already reeling from the 38% annual budget cut, part of a government-wide austerity programme, which targeted science in particular. They took to the streets in weekly demonstrations, and have since launched an online petition, which has been signed by more than 7,000 scientists around the world, including eight Nobel laureates and five recipients of the mathematics Fields Medal. Peter Gruss, president of Germany's Max Planck Society, has written directly to the BAS president offering moral support.

"I worry most about keeping our young scientists," says Tsvetan Dachev, a space scientist at the BAS Space and Solar-Terrestrial Research Institute in Sofia. "They earn just €175 [US\$227] a month, not enough to live on." Last year, 670 scientists left the academy and were not replaced — 460 retired and most



"Winning this battle was tougher than any chemical synthesis I have done."
Daniel Funeriu

R. CHIRITA/MEDIAFAX FOTO

D. DILKOFF/AFP/GETTY

of the rest were young scientists who no longer see a future there.

Things are not much better at Bulgaria's universities, which will see an overall budget cut of 21% to €170 million this year compared with 2009. The large and research-intensive University of Sofia — which alone generates 30% of Bulgaria's scientific papers and patents — has to cope with a 23% cut in its 2009 budget to €18 million. "This is creating tensions between faculties, which are now fighting over tightened resources, and it's not pleasant," says Georgi Vayssilov, a chemist at Sofia. "Faculties such as chemistry and physics need more money for teaching, but our income is based on student number, so we are suffering a lot."

Vayssilov, like most of his colleagues, is also concerned about a law approved last month that allows universities to make their own rules for academic appointments and offers no guidelines or criteria. Consequently, one university might only appoint professors with strong publication lists, he says, whereas another might decide a couple of negligible publications would suffice, opening the door to cronyism.

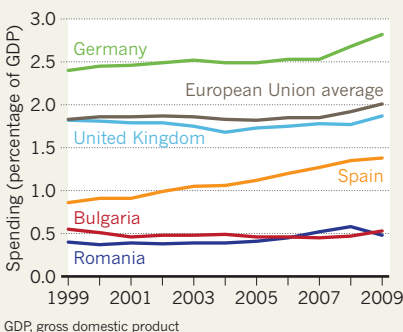
Meanwhile, the country's granting agency, the National Science Fund, put out no general calls for proposals last year, and funding of some ongoing projects has been stopped or reduced. No new calls are expected this year.

TOUGH BATTLE

Until recently, scientists in Romania were in similar despair. In 2008, the country slipped into recession, and the unstable coalition government slashed the 2009 research budget by nearly 50%. But the next coalition government, which came into power at the end of that year, took a different view: a huge budget hike this

THE SPENDING GAP

Research spending in both Romania and Bulgaria lags far behind that of other European nations. After a dip in 2009, Romania now looks set to boost investment in science.



year comes close to returning research spending to 2008 levels. Unlike Bulgaria, Romania is continuing to bolster its research infrastructure, using EU structural funds for regional development.

Crucially, the country is introducing legislation to ensure the new funding is spent well. Research and education minister Daniel Funeriu, a former research chemist who spent most of his adult life abroad, championed an extensive and detailed education law that brings nearly all aspects of university governance, academic hiring practices, evaluation and guidelines for good scientific practice — complete with penalties for academic misdemeanours — into line with other European countries.

The law expressly forbids nepotistic appointments, and also bars rectors from concurrently holding other positions of power that may pose a conflict of

➔ NATURE.COM
For a Q&A with
Daniel Funeriu see:
go.nature.com/f8uyn2

interest, such as being a member of parliament. "If these things didn't happen, we wouldn't have had to write them into the law," says Funeriu.

The draft law met huge resistance in parliament, in large part owing to the influence of several members who are also university rectors. The parliamentary education committee was particularly offended by a paragraph stating that the first round of quality assessment for universities must be done by foreign scientists. In the end, the government decided to force the bill through parliament with a vote of confidence, bypassing some of the parliamentary procedures that were blocking it. After being endorsed by the country's constitutional court on 4 January, the bill became law.

The government is now launching schemes to bring Romanian science into the international arena. In 2010 the country became a full member of CERN, the European particle-physics facility near Geneva, Switzerland, and also became a founding member of two major European facilities: the Extreme Light Infrastructure, which will develop the world's most intense lasers to study matter at high energies, and the particle accelerator Facility for Antiproton and Ion Research in Darmstadt, Germany. This year it is also planning to launch a twinning scheme, which would allow scientists working abroad to set up a parallel lab in Romania. That way they could contribute to science in Romania without having to risk leaving a good job elsewhere; and local staff would benefit from greater contact with international science.

Funeriu says that it has been a hard fight to get to this stage, but he is now optimistic for the future of Romanian science. "Winning this battle was tougher than any chemical synthesis I have done," he says. ■

SOURCE: EUROSTAT

PLANETARY SCIENCE

Space scope finds scorched super-Earth

Kepler's latest discovery is step closer to a true Earth analogue.

BY ADAM MANN

Kepler, the space telescope considered most likely over the next few years to identify an Earth-like planet orbiting another star, has struck solid ground, mission scientists say.

Most of the hundreds of extrasolar planets discovered to date — including the eight previously reported by Kepler — are at least as large

as Neptune and are mainly gas giants. But astronomers hope that the mission will eventually yield a trove of terrestrial planets, including some orbiting their stars at a distance that would allow their surfaces to host liquid water, and possibly life.

Now the team is a step nearer to that goal. It has found a Sun-like star 173 parsecs away — relatively close by galactic standards — harbouring a planet that is just 40% larger than Earth,

although it is uninhabitably hot. Supporting measurements of the planet's mass, gathered at the Keck Observatory on Mauna Kea, Hawaii, suggest that the planet's density is 1.6 times that of Earth, implying a rocky composition. Although other telescopes have already spotted a handful of similar 'super-Earths', the first confirmed rocky planet to fall within Kepler's sights is an encouraging sign for the mission.

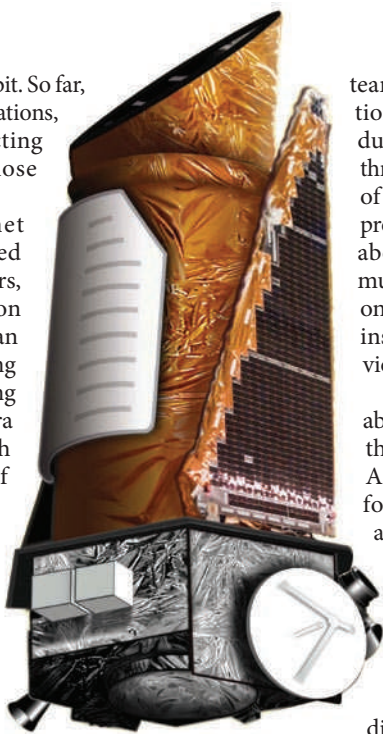
"It's a milestone," says co-investigator Natalie Batalha, an astronomer at San José State University in California, who unveiled news of the discovery during a talk on 10 January at the American Astronomical Society's annual meeting in Seattle, Washington.

Launched by NASA in 2009, Kepler's strategy is to point at a patch of sky in the plane of the Milky Way, where it can continuously monitor more than 100,000 stars, and to look for slight, periodic changes in their luminosity. Such dips in brightness often indicate the presence of planets repeatedly crossing in ►

► front of the stars they orbit. So far, after barely a year of observations, Kepler is limited to detecting short-period planets, those close to their stars.

The new-found planet orbits its star — designated Kepler 10 — every 20 hours, at a distance of 2.5 million kilometres. This is less than 2% of the distance separating Earth and the Sun, ensuring that the planet is more ‘terra lava’ than terra firma, with a surface temperature of 1,800 °C, hot enough to melt silicate rock.

Astronomers at Keck were able to monitor how the planet tugs on the star, giving them its mass. Meanwhile, because the star is nearer than most of the stars in Kepler’s search field, the Kepler



The Kepler craft was launched in 2009.

team could record tiny variations in the star’s luminosity due to sound waves rippling through its interior. The timing of these acoustic oscillations provided detailed information about the star’s dimensions, much as the sound of a string on a bow reveals whether the instrument being played is a violin or a cello.

“Everything we measure about the planet is relative to the parent star,” says Batalha. Armed with a precise figure for the star’s diameter, she and her colleagues were able to nail down the planet’s dimensions with similar precision.

The planet’s close orbit is similar to that of another rocky exoplanet, CoRoT-7b, discovered in 2009, says Greg Laughlin, an astronomer at

the University of California, Santa Cruz, who was not involved in the work. He says the result shows how far astronomers’ techniques for

The first confirmed rocky planet to fall within Kepler’s sights is an encouraging sign for the mission.

finding exoplanets of Earth-like proportions have advanced in recent years. “No one would have thought a decade ago that we’d be finding these ultra-hot rocky super-Earths,” says

Laughlin. Their existence suggests that cooler, smaller worlds await discovery as Kepler’s monitoring time increases and it can identify planets in longer, more distant orbits, he adds.

In February, the Kepler mission is expected to release data on 400 additional planetary candidates. The team is excited, says Batalha, because it hopes that Kepler’s growing number of detections will offer the first concrete estimates of the frequency of extrasolar planets — rocky and otherwise. “So far, it’s been a deluge,” says Batalha. “And we look forward to finding a lot more planets.” ■

AGRICULTURE

Drought-tolerant maize gets US debut

Seed companies race to tap multibillion-dollar market.

BY JEFF TOLLEFSON

When the planting season arrives later this year, farmers in the United States will have a new way to safeguard their crops from drought. Last week, DuPont subsidiary Pioneer Hi-Bred International, headquartered in Johnston, Iowa, announced plans to release a series of hybrid maize (corn) strains that can flourish with less water. The seeds will compete with another maize strain unveiled last July by Swiss agribusiness Syngenta. Both companies used conventional breeding rather than genetic engineering to produce their seeds.

Pioneer says that field studies show its new hybrids will increase maize yields by 5% in water-limited environments, such as the western states of the intensively agricultural Corn Belt region. That compares with the 15% yield gain promised by Syngenta for its maize. Both companies, as well as seed firm Monsanto, based in St Louis, Missouri, are also working on transgenic maize varieties, hoping to tap into a multibillion-dollar market (see *Nature* 466, 548–551; 2010).

In theory, drought-tolerant varieties could fill the gap left in maize supplies in recent years as stocks have been diverted for ethanol production. But not everybody is convinced that these crops will make a big difference. “It’s good news, but it’s not great news,” says David Zilberman, an agricultural economist at the University of California, Berkeley. No crop will survive a severe drought, he says, and other factors such as nutrient availability and soil quality are at play during water shortages, which tend to be more frequent but less severe than droughts. “It will be useful for a small number of really important areas,” Zilberman says, “but my feeling is that people expect altogether too much from drought tolerance.”

Creating drought-tolerant plants has proved to be a difficult challenge for plant breeders. Whereas resistance to a particular herbicide might be pinned down to one gene, the response to drought plays out across the genome. A plant’s resilience also depends on when drought strikes — early or late in the life cycle, for instance — and on which nutrients

are available. Moreover, varieties that perform better when thirsty often underperform when water is plentiful. The industry researchers identified thousands of genes associated with drought tolerance, then incorporated them into their hybrids through conventional breeding. “All of this technology has just come together beautifully in the past ten years,” says Jeff Schussler, senior research manager at Pioneer.

The new varieties will be marketed only in the United States, but efforts to develop drought-tolerant maize are also under way in Africa and Asia. Working with the Bill & Melinda Gates Foundation and other donors, international agricultural research centres have already introduced dozens of conventional drought-tolerant varieties into 13 of the most important maize-producing countries in sub-Saharan Africa. These new varieties now make up 15–25% of the 19 million hectares of maize planted in these countries each year, according to Marianne Bänziger, deputy director-general for research and partnerships at the International Maize and Wheat Improvement Center (CIMMYT) in Mexico City. A recent study by CIMMYT and its partners suggests that the new varieties could increase yields by 10–34%.

The big seed companies are also contributing expertise. Last year, South Africa was home to the first field trial for a transgenic drought-tolerant maize crop, incorporating a gene from the variety currently under development by Monsanto. Meanwhile, CIMMYT recently partnered with the Syngenta Foundation for Sustainable Agriculture, based in Basel, Switzerland, to expand its drought-tolerant maize programme. “Within five years, we hope to have crops in Asia,” Bänziger says. ■

► **NATURE.COM**
For *Nature’s* recent food special, see:
nature.com/food

ENVIRONMENTAL SCIENCE

US science academy report calls for 'bipolar' research

Arctic and Antarctic scientists are urged to share data, ideas and infrastructure.

BY JANE QIU IN PALMER, ANTARCTICA

The Arctic and Antarctic share more than frosty temperatures. Both are experiencing rapid environmental change, and provide sensitive bellwethers for the effects of global warming. Yet scientists studying the ends of the Earth are poles apart in more than just geographical terms, which is hampering research and slowing efforts to understand how ecosystems are responding to climate change.

The warning comes from a strategic report commissioned by the Polar Research Board, a subgroup of the US National Academy of Sciences (NAS), which recommends that the community reshape itself to take a 'bipolar' approach, sharing resources and initiating more holistic research projects to identify common patterns of change at both poles. The report, 'Frontiers in understanding climate change and polar ecosystems', is due for release in the coming months. Its authors hope that it will encourage granting bodies such as the US National Science Foundation to promote more integrated polar research.

Both the Arctic and Antarctic are "highly coupled systems" where changes to land, ocean or atmosphere can have enormous effects on the region as a whole, explains John Prisco, an ecologist at Montana State University in Bozeman, who has worked at both poles during the past 25 years, and is a co-chair of the NAS study. Marine ecosystems at both poles depend on the existence of sea ice, but current models cannot accurately predict recent changes in sea-ice cover. Sharing data on Arctic and Antarctic sea-ice decline could highlight common responses to climate change and help to develop better predictive models of its consequences, the report will argue.

In the Arctic, where the rise in air temperature has been almost twice the global average



Antarctica's Adélie penguins are shifting further south as warmer temperatures affect the food chain.

over the past few decades (go.nature.com/xph8q6), permafrost is thawing on a large scale. This is driving changes in vegetation, such as a shift from moss to shrubs, and accelerating carbon release from the ecosystem. In the western Antarctic Peninsula, where mid-winter temperatures have increased by about 6°C and the sea-ice season has shrunk by nearly 90 days during the past 50 years (O. Schofield *et al. Science* **328**, 1520–1523; 2010), there are signs that the ecosystem "may be approaching what could be an absolutely catastrophic change", says William Fraser, president of the Polar Oceans Research Group in Sheridan, Montana.

Understanding the tipping points beyond which polar ecosystems would change rapidly and irreversibly is "a gigantic priority", says report co-author Hugh Ducklow, an oceanographer at the Marine Biological Laboratory in Woods Hole, Massachusetts.

Beyond data-sharing, the report urges researchers to take practical steps. To make results from different polar regions more comparable, it says, experimental approaches and measurements of factors such as sea-ice extent or population dynamics should be standardized. Researchers could share expertise about the best way to use infrastructure. They could even pool expensive equipment such as sensors, small submarines, helicopters and icebreakers, shipping it from pole to pole as needed, says Oscar Schofield, an oceanographer at Rutgers University in New Brunswick, New Jersey.

The need for such an approach is urgent, says Fraser. "The ecosystem is being stressed to the point where something has got to give." ■

Jane Qiu was in Antarctica as part of the 2010 Marine Biological Laboratory's Logan Science Journalism Fellowship.



LATEST NEWS



Neuroscientist quits as fraud investigation rocks Danish university
go.nature.com/k1pglr

MORE STORIES

- Plasma jets key to enduring solar mystery go.nature.com/egqta
- Glassy metal set to rival steel go.nature.com/xgxqtn
- Why dire climate warnings boost scepticism go.nature.com/3yvdja
- Tears are a turn-off go.nature.com/9o2v6t

VIDEO



X-rated worm movies reveal sex secrets
go.nature.com/8vjwv

CONSERVATION

Plans for marine protection highlight science gap

Researchers are scrambling to understand how best to deploy conservation zones.

BY DANIEL CRESSEY

Facing a host of threats including fishery devastation and the destruction of coral reefs, conservationists have increasingly pinned their hopes on marine protected areas (MPAs). More than 5,000 of these sanctuaries, where fishing is controlled to limit its effect on biodiversity, have been set up, mainly in coastal zones. They range in size from less than 10,000 square metres to the vast Phoenix Islands area, part of the Republic of Kiribati in the Pacific Ocean, which tops 400,000 square kilometres.

But, in the rush to safeguard marine ecosystems, there has been little scientific assessment of how to create a successful MPA. With a new wave of MPAs expected to be created in deep-ocean regions in the next few years, scientists are eager to understand how factors such as size and siting can determine a protected area's success or failure.

Tundi Agardy, an environmental consultant based in Colrain, Massachusetts, is the lead author of a paper¹ published in *Marine Policy*, which warns of a "blind faith" in the ability of MPAs to stem biodiversity loss. She told *Nature* that she can name only "a handful" of areas that actually work as advertised. Her paper identifies five possible shortcomings in MPAs: many are too small to be effective; they may simply drive fishing into other areas; they create an illusion of protection when none is actually occurring; many are poorly planned or managed; and they can fail all too easily because of environmental degradation of waters just outside the protected area.

"I'd venture a guess that a majority of the world's several thousand MPAs have one flaw or another relating to the five categories we describe," says Agardy. For example, an MPA created to protect the vaquita (*Phocoena sinus*), a small porpoise found only in the Gulf of California, actually missed a sizeable proportion of the species' core range. The animal's numbers have continued to decline and it is now the most endangered marine mammal in the world².

"We still need a lot of knowledge to really understand how MPAs work exactly," says Frederic Vandeperre, a marine biologist at the University of the Azores in Horta who last month published an analysis³ of seven southern European MPAs. The study showed that these MPAs can benefit fisheries in adjacent waters,



P. NICKLEN/NAT. GEOGR./GETTY

The reefs of the Phoenix Islands in Kiribati are part of the world's largest marine protected area.

but that the degree of the effect depends heavily on the size of the area and the quality of its management. Vandeperre says that each MPA needs a unique design, depending on its goals. For example, those that explicitly aim to safeguard fishing yields need to cover a larger area.

INTERNATIONAL WATERS

Conservationists should approach the design and siting of an MPA as an experiment, he says. "We should maybe create MPAs with different structures, different designs in a controlled way, to be really like an experiment so we can figure out which elements are crucial." This could include variations in size, location, management strategy, monitoring and proximity to other MPAs.

Understanding the best way to create MPAs is about to become much more important. The 2002 World Summit on Sustainable Development set ambitious, internationally agreed targets to establish extensive networks of MPAs around the world by 2012.

This requires the creation of more MPAs outside national boundaries in the high seas, where still less is known about how to make them work. "We have almost no experience of applying marine protected areas to high-seas ecosystems," says Alex Rogers, a conservation ecologist at the University of Oxford, UK. "We don't really know where to put them. We suspect that simply by placing them in places that

are particularly sensitive for species we may be able to derive a considerable management benefit, but it's very, very early days at the moment." Rogers is organizing a conference at the Zoological Society of London next month to discuss the design of high-seas protected areas, along with the complex legal and political issues that surround them.

Still, some studies are starting to give clear pointers on the best way to position both coastal and high-seas MPAs. Last month, Mark Christie of Oregon State University in Corvallis and his team published an analysis⁴ showing that fish larvae — those of the yellow tang (*Zebrasoma flavescens*) — were successfully dispersing from an MPA to sites up to 180 kilometres away.

"Now we are able to show the larvae can drift to sites outside the MPA and essentially reseed fish stocks significant distances away," says Christie. The result means that by combining information about ocean currents with the genetics of larvae captured from the seas, researchers can identify from where the larvae came. That could help pinpoint — and protect — the most important spawning areas for species such as Pacific bluefin tuna (*Thunnus orientalis*), says Rogers. ■

1. Agardy, T., Notarbartolo di Sciara, G. & Christie, P. *Mar. Policy* **35**, 226–232 (2011).
2. Dalton, R. *Nature* **465**, 674–675 (2010).
3. Vandeperre, F. et al. *Fish Fish.* doi:10.1111/j.1467-2979.2010.00401.x (2010).
4. Christie, R. et al. *PLoS One* **5**, e15715 (2010).



THOUGHT EXPERIMENT

Japanese hospitals are using near-infrared imaging to help diagnose psychiatric disorders. But critics are not sure the technique is ready for the clinic.

BY DAVID CYRANOSKI

In a room full of psychiatrists in downtown Tokyo, I prepare to have my mental health assessed. No probing questions are asked. Instead, I don an odd type of swimming cap, criss-crossed with cables and studded with red and blue knobs. At the flick of a switch, the 17 red knobs send infrared light 2 to 3 centimetres into my brain, where it is absorbed or scattered by neurons. Photoreceptors in the 16 blue knobs retrieve whatever light bounces back to the surface. Buried in the signals, say the researchers operating the system, are clues that can distinguish depression, bipolar disorder, schizophrenia and a normal state of mind.

More than 1,000 people have already been subjects of the technique, called near-infrared spectroscopy (NIRS) and developed by Masato Fukuda, a psychiatrist and neuroscientist at Gunma University Hospital in Maebashi, and the Hitachi Medical Corporation in Tokyo. Most of those were research subjects. But since April 2009, when NIRS was approved by the health ministry as an “advanced medical technology” to assist psychiatric diagnoses, more than 300 people have paid ¥13,000 (US\$160) out of their own pocket to access the technique. The University of Tokyo Hospital, one of eight leading Japanese research hospitals now offering NIRS diagnostic neuroimaging, found demand for it to be so high that the hospital stopped taking appointments twice. Gunma University Hospital is fully booked to the end of March. “We’ve been overwhelmed by requests,” says Fukuda.

The appeal of NIRS is its promise of fast, clear-cut diagnoses of psychiatric conditions which, with their messily overlapping symptoms, are frequently diagnosed wrongly or not diagnosed at all. US studies, for example, found that some 70% of bipolar patients were initially misdiagnosed^{1,2}. As for patients, says Fukuda, “They want some kind of hard evidence,” especially when they have to explain absences from work.

NIRS could offer an objective measure of mental health reliable and convenient enough for routine use in the clinic. Fukuda says that it can help point to a diagnosis much like a chest X-ray might be used to help diagnose pneumonia or an electrocardiogram to define a heart problem. Aside from Fukuda and a group of doctors in Japan, however, few scientists are persuaded.

➔ NATURE.COM
More on
schizophrenia at
nature.com/schizophrenia

Critics charge that the studies so far have been too few, too small and too weakly designed to warrant the technique’s clinical use. “It’s attractive as a research topic, but the data are not convincing enough,” says Masahiko Haruno, a neuroscientist at Tamagawa University in Tokyo. John Sweeney, a neuroscientist at the University of Illinois, Chicago, who has spent two decades looking for connections between various brain-monitoring techniques and diseases such as schizophrenia, says that “none has ever been validated to anyone’s satisfaction”. And NIRS is the least developed of them all, he says, calling it “the thinnest of ice to be treading upon. We are nowhere near ready to tell patients and families that they should have these kinds of tests.”

Y. AOYAMA

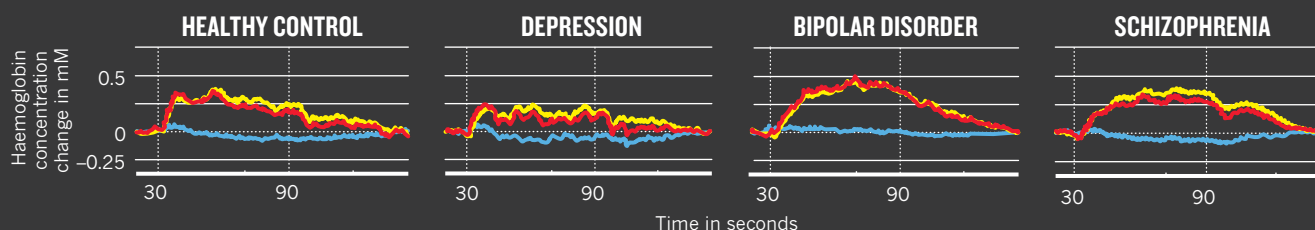
NEW KID ON THE BLOCK

Sporting the knobby cap, I stare at a screen showing Japanese phonetic characters and say aloud words beginning with each sound. The words don’t come readily, particularly in front of an audience of psychiatrists and neuroscientists. Coming from all over Japan, they meet every month to discuss NIRS results and strategies.

Having entered the research scene some 15 years ago, NIRS is relatively new compared with functional magnetic resonance imaging (fMRI) or electroencephalography (EEG), but in Japan it has raced ahead to the clinic. (The two biggest suppliers of NIRS analytic devices — Hitachi and Kyoto-based Shimadzu — are both based in Japan and the country accounts for two-thirds of publications using NIRS analysis.) The technique takes advantage of the fact that compared with constituents of other tissues, haemoglobin in blood absorbs more light in the near-infrared spectrum. Blood flow to a particular brain region increases when neurons there are active. So monitoring the changes in haemoglobin concentration gives a site-specific read on blood flow and thus on neuronal activity³. Fukuda’s NIRS device focuses on the prefrontal cortex and temporal cortex, regions that are implicated in many of the symptoms seen in psychiatric disorders; the signature pattern of blood flow associated with each disorder is used to help diagnose it. NIRS lacks the precision and depth of fMRI, which can pinpoint changes in blood flow throughout the brain and with much greater spatial resolution.

TRACES OF TROUBLED MINDS

Characteristic patterns of cortical blood flow measured by near-infrared spectroscopy.



Left: Headgear for probing blood flow in the brain with near-infrared light.

But NIRS is relatively cheap and mobile, and subjects can sit upright without having to endure a spell in the large, loud and sometimes nerve-racking tube of an fMRI machine. This means that NIRS is easier to use on fidgety subjects such as children, and people with psychotic conditions or anxiety. The advantages have made infrared imaging increasingly popular with brain researchers worldwide. Devices from the largest maker in the United States, NIRx Medical Technologies of Glen Head, New York, are being used to study areas ranging from autism to brain-computer interfaces. Hitachi now offers a stripped-down version that allows the brains of four people interacting in a room to be analysed wirelessly.

Fukuda, though, has focused on applying the technology to diagnoses. Lean and grey-haired, he speaks thoughtfully and is confident in the technique and its potential to help people. He started using a basic NIRS device in 1997, when he was an associate professor at the University of Tokyo already working with EEG. Since then, he has been measuring the brain activity of people with a variety of disorders and has reported that those with depression, bipolar disorder or schizophrenia have, when averaged across groups of 10–20, a characteristic pattern of brain activation^{4,5}. Fukuda boils the data down to graphs describing activity in the prefrontal and temporal cortices for the first 60 seconds or so of each task (see ‘Traces of troubled minds’). He says that the NIRS test on its own classifies patients correctly 80% of the time.

These studies have not convinced other neuroscientists. Haruno says the patient numbers in the published studies are “far too small” to distinguish patterns, and that even if such patterns are found when signals are averaged across groups, this does not mean that one person’s pattern can be used to assign them to a group. “What does that mean for an individual patient? It’s very misleading,” Haruno says.

Even Fukuda’s collaborator, Andreas Fallgatter of the University of Tübingen in Germany, who has used NIRS for 14 years on about 1,000 patients and is now repeating Fukuda’s language test in German, says “NIRS is still a research method.” Still, he says, “Obviously, Dr Fukuda could successfully convince the Japanese authorities.”

That approval came via a fast-track process instituted by the national health ministry’s Advanced Medical Technology programme in 2005 in an attempt to spur the development of biomedical technologies. Teruhiko Higuchi, a clinician and researcher specializing in depression and president of the National Center of Neurology and Psychiatry, led the evaluation of NIRS. He concluded that the technology was safe, effective and fast, and could help to make critical distinctions between different mental states (major depression, bipolar disorder or schizophrenia) at an early stage, when used with other diagnostic techniques. “It is, in the end, only to assist diagnosis,” he told the members of the evaluation committee, according to meeting minutes posted online. Higuchi’s centre now offers the technique. Other committee members raised concerns about the small numbers of patients in the studies, and the fact that some were receiving drugs, but they did not object to its approval.

Fukuda says that a larger study involving more than 500 patients will be submitted for publication soon and will answer many of his critics.

He says that doubling the number of knobs and other methodological modifications reveal a much sharper distinction between the conditions, and that controlling for medicated versus non-medicated patients showed that drugs do not obscure a patient’s NIRS profile. He acknowledges the validity of criticism about using group averages to diagnose individuals: “Strictly speaking, this criticism is right.” But he says the same is true for many other measurements, such as electrocardiograms and EEG, which vary from one individual to the next and thus require interpretation, but can still be clinically useful. “Clinical diagnosis and NIRS examination are complementary to each other,” he says. “We stress this complementary nature to all the patients.”

But in at least some cases, NIRS seems to take the lead in diagnosis. For example, when Fukuda calculates his success rates, NIRS results that match the clinical diagnosis are considered a success. If the results don’t match, Fukuda says he will ask the patient and patient’s family “repeatedly” whether they might have missed something — for example, whether a depressed patient whose NIRS examination suggests schizophrenia might have forgotten to mention that he was experiencing hallucinations. Andreas Meyer-Lindenberg, an expert in neuroimaging and mental health at the University of Heidelberg, Germany, says that studies of patients without an existing diagnosis or psychiatric medication would be more persuasive. “You would need a sample of unclear cases, as you would get in the clinic, classify them and then ascertain their diagnosis by following them up.”

Fukuda and his colleagues are already moving on to NIRS studies that might aid diagnoses of a range of disorders, including those centring on panic, attention deficit and post-traumatic stress.

My own NIRS results, however, are short on clarity. Within 15 minutes, including the tests and a quick computer analysis, Fukuda is able to look at my traces and deliver a diagnosis: normal. When I later compared them to the patterns published in the literature, however, my trace seems to describe a brain somewhere between normal and bipolar.

Later, Fukuda says that my pattern is not a normal ‘normal’ NIRS trace, perhaps because the observers in the room made me hesitate to speak. He also says that a subset of healthy subjects has the pattern that I do in the frontal lobe and that data he measured from the temporal lobe helped him reach a diagnosis. However they are reached, I suppose I should be happy about my results. ■ [SEE EDITORIAL P.132](#)

David Cyranoski is Nature’s Asia-Pacific correspondent.

1. Lish, J. D. Dime-Meenan, S., Whybrow, P. C., Price, R. A. & Hirschfeld, R. M. *J. Affect. Disord.* **31**, 281–294 (1994).
2. Hirschfeld, R. M., Lewis, L. & Vornik, L. A. *J. Clin. Psychiatry* **64**, 161–174 (2003).
3. Maki, A. *et al. Med. Phys.* **22**, 1997–2005 (1995).
4. Suto, T., Fukuda, M., Ito, M., Uehara, T. & Mikuni, M. *Biol. Psychiatry* **55**, 501–511 (2004).
5. Kameyama, M. *et al. NeuroImage* **29**, 172–184 (2006).

SOURCE: M. FUKUDA *ET AL.* IN: FUKUDA M. (ED) *NIRS AND PSYCHIATRIC DISORDERS* (NAKAYAMA SHOTEN, 2009)



THE END OF THE WILD

BY EMMA MARRIS

Climate change means that national parks of the future won't look like the parks of the past. So what should they look like?

Imagine Montana's Glacier National Park without glaciers; California's Joshua Tree National Park with no Joshua trees; or the state's Sequoia National Park with no sequoias. In 50 years' time, climate change will have altered some US parks so profoundly that their very names will be anachronisms.

Jon Jarvis, who became director of the US National Park Service in 2009, has called climate change "the greatest threat to the integrity of our national parks that we have ever experienced". The sentiment represents a dramatic shift from the position held during the Bush administration, when officials refused to fully acknowledge

the existence of climate change. Now, park managers in the United States and around the world are working with researchers to map how the landscapes they care for might change. And they are coming to terms with the idea that the historical remit of most parks systems — to preserve a piece of land in its 'natural' state — is untenable. "You can't fight the climate," says

NATURE.COM
To hear more on
options for managing
Yellowstone, visit:
go.nature.com/ya6k7y

Ken Aho, an ecologist at Idaho State University in Pocatello, who studies non-native species at Yellowstone National Park in Wyoming. "Eventually

The Grand Canyon of Yellowstone by Thomas Moran (1872).

you have to throw up your hands,” he says.

Nowhere is attachment to historical fidelity more pronounced than at Yellowstone, the first US national park and the best example of the park as a landscape seemingly unchanged by the passage of time. Visiting it, one crosses paths with bison and wolves. It is not hard to imagine a party of explorers coming around the next bend.

Much of Yellowstone's 900,000 hectares are high plateau, crossed by rivers and dotted with geothermal basins featuring pools and geysers. The park was created in 1872, to protect the geological wonders and safeguard a wild landscape emblematic of the American west. Since 1916, Yellowstone and the nation's other parks and monuments have been run by the National Park Service, which has aimed to preserve the land in its 'natural' condition — typically meaning how it looked before white people arrived. In the words of the influential *Leopold Report* of 1963: “A national park should represent a vignette of primitive America” (A. S. Leopold *et al.* *Wildlife Management in the National Parks* Advisory Board on Wildlife Management; 1963.)

But Yellowstone can no longer be kept as it was in 1872, if that were ever possible. Climate change has already begun to transform the park (see ‘Under attack’). The Yellowstone area has been plagued by tree pests, abetted by warmer temperatures. Fires are expected to become more frequent, animal populations are shifting and the landscape and ecology are being reshaped. Similar stories are playing out at all of the nation's parks, and the Park Service is beginning to react. In September 2010, it released a *Climate Change Response Strategy*, which includes sections on science, adaptation, mitigation and communication (National Park Service Climate Change Response Program; 2010). The report hints that climate change may force the Park Service to change the way it defines its mission. Maintaining a natural state can no longer be the goal, or important tools for adaptation, such as moving species or selective breeding, would be forbidden. And besides, that battle may already be lost.

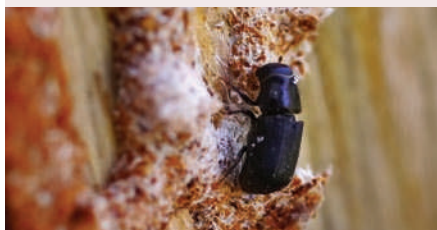
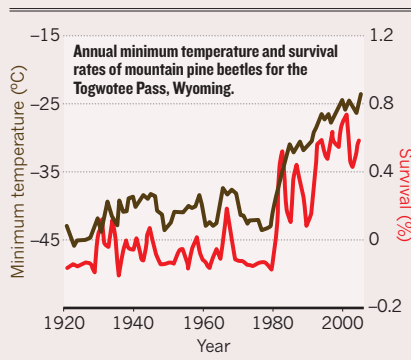
ENCROACHING DAMAGE

A drive through the greater Yellowstone area reveals changes to the landscape. Most striking are the acres of trees standing dead, killed by an insect the size of a grain of rice. Mountain pine beetles, native parasites, burrow into and reproduce in the living wood of the trees. Winter temperatures of -40°C kill the beetles, keeping their numbers down. But warm winters in the past ten years have allowed them to proliferate. More than half of the greater Yellowstone ecosystem's conifer forest has seen pine-beetle damage, and 10% of the forest has a ‘high severity’ infestation, in which more than 40% of trees are lost. Throughout the western United States,



UNDER ATTACK

Climate change may increase the frequency of intense fires similar to the historic 1988 blaze (above) and may promote destruction of the trees by mountain pine beetles (bottom), which flourish in rising temperatures (graph).



the Park Service has used insecticides to protect some trees, and removed a few dead ones. Mostly, however, it can only watch as an orange wave of dying trees ripples through the forest.

The dried-out remains left in the beetles' wake look like an invitation to intense forest fires that would reshape the landscape even more, but that is not necessarily the case. Monica Turner, an ecologist at the University of Wisconsin–Madison, has been modelling big blazes at Yellowstone, such as the inferno of 1988 that burned 321,000 hectares of forest. She and her colleagues have found that such fires burn in the living crowns of trees and don't rely on dead wood; furthermore, pine-beetle attacks kill trees, decreasing the amount of wood in an area and creating gaps that can stop a fire.

Even so, Turner's group expects that the big fires that typically occur every few hundred years might happen more frequently in the future without the beetles' help, thanks to increased temperatures and changes in precipitation levels. That increased frequency might turn some parts of Yellowstone into a landscape of young forests, with less capacity than more mature ones to absorb carbon dioxide from the atmosphere. This could potentially turn the park at large from a carbon sink into a carbon source.

But forecasting the future is complicated. Pines that don't get caught in a blaze may actually grow better in a warmer future; growth of lodgepole pines is limited by the length of the growing season and by cold temperatures. Meanwhile, species such as the larch, currently confined to lower altitudes than those at Yellowstone, may move up as temperatures rise.

Other familiar tree species might go locally extinct. Whitebark pine, a high-altitude species, is facing a three-way attack from pine beetles, a fungal disease called blister rust and climate change. As Yellowstone's peaks heat up, the tree's range is expected to move upslope, but the higher it goes the less area it will have. “The future of whitebark pine in Yellowstone is questionable,” says Turner.

Ecosystems won't move predictably. “Migration in response to climate change can often be extremely messy,” says Stephen Gray, a climatologist at the University of Wyoming in Laramie, who is among those working on scaling down global climate models to forecast changes at the level of individual parks.

As ecologists scramble to predict changes, park managers are gearing up for a new management style, which will have to include at least one of two approaches traditionally anathema to the profession: letting things change, or intervening aggressively to keep them the same. In many cases, choosing between these strategies will be the challenge.

If managers choose the former, they will need to create an environment conducive to change. For example, many conservation biologists argue for creating and maintaining corridors that connect parks to other natural areas. The bigger the connected area, the more room

SPECIES MANAGEMENT

FOUR SEASONS OF WORRY FOR THE GRIZZLY

plants and animals will have to move and the larger the gene pools available for adaptation.

Wolverines are one species that might benefit. Maps of future climates suggest that by 2040, lower-elevation parts of their current range may no longer have the deep spring snow that wolverines need to make their dens. So managers should perhaps focus on increasing the quality and connectivity of land in the Colorado Rocky Mountains, California's Sierra Nevada and parts of Wyoming and Utah to make sure that the wolverine population remains large enough to be genetically healthy.

To maintain corridors in areas that fall outside national-park remits, the Park Service is participating in Landscape Conservation Cooperatives, an initiative begun in September 2009. These clubs of federal and state agencies, conservation organizations, university scientists and other interested parties focus on managing huge chunks of land as units, rather than having every park, forest and piece of private land managed independently.

A FIRM HAND

But tough choices loom in terms of how much to meddle. Aggressive intervention might be needed to conserve some of Yellowstone's larger mammals (see 'Four seasons of worry for the grizzly') and its iconic trees. To preserve the whitebark pine, some scientists from various agencies have begun identifying trees that are resistant to blister rust and collecting seeds to breed rust-resistant trees. They've also begun planting whitebarks in newly burned areas. They plan to use insecticides to protect individual trees from pine beetles, prune off infected branches and thin the vegetation around the trees to give them a competitive advantage. They also plan to stop fires from burning rust-resistant or particularly genetically diverse stands of whitebarks. Some of these planned actions sound a lot like landscape gardening.

The alternative approach would be to let the whitebarks die out in Yellowstone, and plant them somewhere where they might flourish in a warmer future. Sally Aitken, a forest geneticist at the University of British Columbia in Vancouver, Canada, has planted whitebarks beyond their current range, in northwestern British Columbia. They seem to like it there.

Jarvis has suggested the possibility of moving species outside their native ranges to give them a better chance of surviving — just not right away. "The big point here is that we are willing to face these questions," he says. "We are not afraid to talk about them."

Other countries have ditched 'naturalness' for goals that encompass a range of acceptable states, and don't rule out aggressive intervention. Parks Canada uses an 'ecological integrity' approach, preserving a configuration of local flora and fauna that is likely to be robust.

At a conference on the Yellowstone ecosystem last October, many speakers talked about managing for 'resiliency' rather than historical

Grizzly bears can take the heat. If temperatures go up by a few degrees in Yellowstone National Park in Wyoming, they will just shed a little hair. But other changes might put the beasts to the test in a warmer future. Their movements tend to revolve around a few food sources spread out over the year — several of which are threatened by climate change and exotic species.

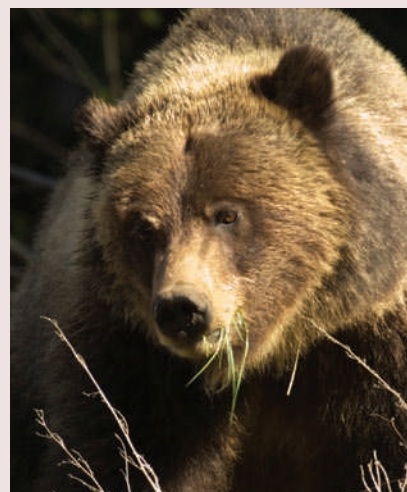
In spring, bears emerge from their dens and line the streams feeding Yellowstone Lake, looking for cutthroat trout that are migrating upstream to spawn. These days, the bears sometimes fish in vain. Introduced lake trout are eating the cutthroat trout, and lake trout stay in the lake to spawn.

In summer, grizzlies gorge on army cutworm moths found on high-altitude rocky slopes in Yellowstone. "A grizzly can eat about half its energy needs for the year in about 30 days of foraging on moths," says Hillary Robinson, an ecologist in Bozeman, Montana. If climate change dries up the wet microhabitats that the moths prefer or reduces the populations of prairie flowers that the caterpillars feed on, Yellowstone's grizzlies will feel the pinch.

By autumn, whitebark-pine seeds, thoughtfully collected into caches by squirrels, are a favourite food for grizzlies. But whitebarks are declining owing to an introduced fungus and an outbreak of mountain pine beetles that has been blamed on climate change. Roughly 82% of whitebark stands in the greater Yellowstone area are dead or dying.

Problems continue in late autumn and winter, when grizzlies that haven't yet gone into hibernation go looking for elk and other ungulates killed by the harsh conditions. If winters mellow, there may be fewer carcasses. This shortage of meat may be behind a trend of more bears leaving the park in winter and feeding on hunters' gut piles in recent years.

As if that weren't enough, Hank Harlow, a bear physiologist at the University of Wyoming in Laramie, is worried that warm



winters will rouse hibernating bears, reducing the time they spend in energy-saving sleep. The energy expended on waking up could eat into the 5 kilograms of fat that females need to birth and nurse a pup during hibernation.

Park managers fear that as their food sources wane, bears will roam into the human environment, looking for hunters' kills, sheep and refuse. Many will be shot; humans are a leading cause of death for the bears. "Everywhere they go outside of public land there is the urban, human interface," says Mark Haroldson, a biologist with the Interagency Grizzly Bear Study Team at the US Geological Survey in Bozeman. "It is a minefield for them."

If climate change and human pressure reduces the food available until the land can no longer support a viable bear population, Yellowstone's managers may have to decide between letting grizzlies go and feeding them — an option at odds with attempts to maintain the park as a wilderness.

"Feeding bears just bothers me," says Leigh Welling, the Climate Change Response manager for the US National Park Service. But, she adds, "I don't think we would close the door on it". **E.M.**

fidelity, promoting an ecosystem's ability to change with the times without changing in character. Rather than trying to sustain a stressed ecosystem that collapses at the first hot summer, the idea is to let things change gracefully.

But the Park Service isn't planning to change its mission any time soon. "I don't think resilience replaces our current management foundation," says Jarvis. "Our goal has never been to freeze [parks] in some kind of stasis."

Resiliency may be achievable for Yellowstone, even if it won't look like it did in 1872. Whitebark

pinus may not make it, but lodgepole pines are very resilient. Bears and wolves are clever. Yellowstone is big and Yellowstone is tough.

"As a scientist who really treasures the region," says Turner, "I believe that Yellowstone will go on in the face of climate change. Yellowstone is very resilient. The 1988 fires are not a catastrophe. Bark beetles are not a catastrophe." ■ **SEE EDITORIAL P.131**

Emma Marris writes for *Nature* from Columbia, Missouri.

COMMENT

MEDICINE A plan to overhaul the disappointing search for biomarkers **p.156**

OCEANS Governance struggles to protect and exploit seas under pressure **p.158**

NEUROSCIENCE Exhibition explores the wonders of the brain **p.161**



COMMUNICATION A call to evaluate the quality of outreach **p.162**



Port-au-Prince, Haiti, 2010.

Corruption kills

On the anniversary of Haiti's devastating quake, **Nicholas Ambraseys** and **Roger Bilham** calculate that 83% of all deaths from building collapse in earthquakes over the past 30 years occurred in countries that are anomalously corrupt.

The six-digit death toll from last year's Haiti earthquake compared with the absence of any fatalities in New Zealand's identical magnitude (7) earthquake was a stark reminder that poor building practices are largely to blame for turning moderate earthquakes into major disasters. Earthquake-resistant construction depends on responsible governance, but its implementation can be undermined by corruption¹⁻⁵ or by poverty, through the use of substandard materials and assembly methods, or through the inappropriate siting of buildings^{6,7}.

The effects of these forces are difficult to tease apart, because the poorest nations are

often also the most corrupt. To try to isolate these influences, we quantified a global relationship between national corruption⁸ and a nation's per capita income⁹. It showed that some nations are more corrupt than anticipated. It is in these countries that about 83% of all deaths from earthquakes in the past three decades have occurred.

The construction industry — currently worth US\$7.5 trillion annually and expected to more than double in the next decade — is recognized as being the most corrupt segment of the global economy¹⁰. Corruption takes the form of bribes to subvert inspection and licensing processes, and of covert

activities that reduce costs and thereby compromise the quality of structures. The assembly of a building, from the pouring of foundations to the final coat of paint, is a process of concealment, a circumstance ideally suited to the omission or dilution of expensive but essential structural components. Reports of the spontaneous collapse of new buildings testify to a lack of construction oversight (for example, Shanghai, 27 June 2009; Delhi, 15 November 2010). During earthquakes, the consequences of decades of shoddy construction are revealed on a catastrophic scale^{1,2,4,11}.

The analysis we present here is a sequel ►

L. ABASS/MINUSTAH/GETTY IMAGES

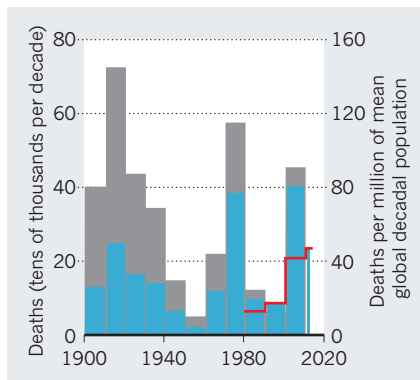


Figure 1 | Earthquake deaths. Despite advances in earthquake engineering, the number of people killed by earthquakes each decade has surged (blue), and the number of deaths as a proportion of global population has not dropped much (dark grey). Many of these deaths can be attributed to building collapse (red).

► to a report on the mitigation of earthquake losses that one of us (N.A.) wrote in 1976 for the United Nations Educational, Scientific and Cultural Organization. The section of that report discussing the causes and effects of a lack of transparency on engineering failures was never published⁵. Following more recent earthquakes in China, Pakistan and Haiti, we felt it was imperative to update and air that discussion.

GROUND RULES

The number of deaths continues to climb despite advances in earthquake-resistant design in the past century (Fig. 1). Averaged over the past decade, the fatality rate is 60,000 a year. This average is dominated by the earthquakes in Indonesia in 2004, Kashmir in 2005, Iran in 2005, China in 2008 and Haiti in 2010. It includes fatalities from building collapse and from secondary causes such as tsunamis, landslides and fire. Since 1980, fatalities from dwelling collapses, for which an absence of earthquake engineering may be held responsible, average 18,300 a year.

The recent increase in earthquake fatality rates might be supposed to arise from population growth, urbanization or industrial development. Indeed, when adjusted for population growth, deaths from earthquakes are loosely linked with average global populations (Fig. 1). So has the application of earthquake-resistant design and construction of dwellings had any effect in reducing fatalities from earthquakes? Yes: significant nation-to-nation variations in the cumulative death toll indicate that the application of resistant engineering clearly benefits earthquake-prone countries that have the wealth and willpower to mandate its use.

Corruption is by nature covert and difficult to quantify. Yet business people or foreign

visitors are frequently willing to disclose its presence on the condition of anonymity. The degree to which corruption is perceived to exist in different countries has been ranked annually⁸ since 1995 by Transparency International, a global civil society organization headquartered in Berlin, using a Corruption Perceptions Index (CPI). The score is determined from an aggregate of 13 polls averaged over 2 years from 10 institutions alert to the frequency and extent of bribes paid within various countries. A CPI score of 0 indicates a highly corrupt nation with zero transparency; a score of 10 indicates an absence of perceived corruption with total transparency. The CPI is less reliable for countries with fewer sources of information¹². We used an average CPI derived from our investigation of long-term fluctuations (Supplementary Fig. 1 — Fig. S1), and its standard deviation (Fig. S2).

Relative wealth is the most obvious parameter that influences a country's corruption. Wealth is frequently attended by a stable constitution conducive to the rule of law. A standard measure that allows comparison of wealth between countries or across economies is the gross national income (GNI) per capita. We chose the World Bank's GNI Atlas method⁹ with data averaged over the period 1960 to 2009 (Figs S3 and S4). A clear correlation exists between a nation's per capita income and the level of corruption (Fig. 2). The most corrupt nations are the poorest (Figs S5 and S6).

For earthquakes of the twentieth century, particularly the first half, it is not always possible to confirm published fatality estimates or to calculate new reliable ones. Previous catalogues characterize uncertainties in fatality counts by listing estimates from

multiple sources uncritically¹³. The weighting of the most reasonable number from these is largely subjective⁴.

We devised a new catalogue by examining original sources such as government reports and aid-agency responses. That said, even for the 2010 Haiti earthquake, reported fatality estimates vary by a factor of three from fewer than 85,000 (an investigative count — probably accurate) to 300,000 (an unsubstantiated guess) (Table S1). Our catalogue distinguishes deaths caused by the collapse of dwellings due to ground shaking from the total number of earthquake-associated deaths, which include those from secondary effects¹⁴ such as aftershocks, landslides, fire and tsunamis.

The number of deaths attributable to the collapse of dwellings is influenced by population density and the vulnerability of building stock in the epicentral region. In the past 30 years, the rapid increase in urban populations, particularly in developing countries, has adversely influenced building quality. The number of fatalities depends on whether an earthquake happens at night or during the day, in the winter or in the summer, in a mountainous region or in a valley, after strong and protracted foreshocks and with or without warning¹⁵. An earthquake occurring on a winter night is likely to kill two to five times more people than one on a summer morning, particularly in a rural region.

GEOLOGY'S ACCOMPLICES

We compared earthquake fatalities from 1980 to 2010 with measures of corruption and wealth. We found, as expected, a direct relationship between poverty and deaths

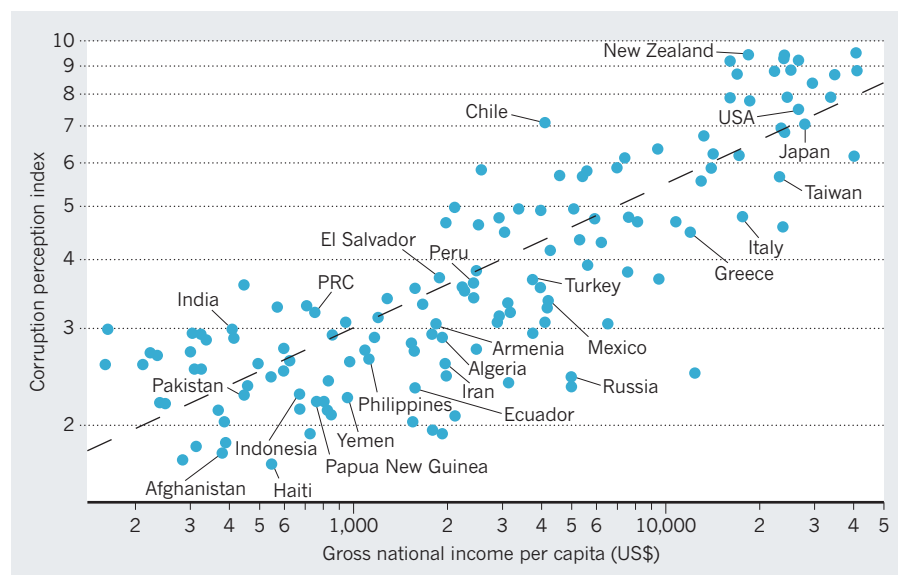


Figure 2 | Cash and corruption. The poorest countries are the most corrupt, but some are more corrupt than others. A weighted regression line (dashed) divides nations that are perceived as more corrupt (below the line) than might be expected from the average income per capita from those that are less corrupt (above the line). Named countries have lost citizens in building collapse caused by earthquakes since 1980.

from earthquakes. Clearly, poverty can lead to the use of unsatisfactory building materials (such as adobe or poor-quality concrete), and to a paucity of education, resulting in ignorance in construction. We also found that corrupt societies have the largest death tolls from earthquakes. For the period 1995 to 2010, when corruption values can be compared directly with earthquake fatalities, we find a quantitative link between the two (Figs S10 and S11). Because the corruption index changed only slowly in this interval (Fig. S1), we assume that CPI values for 1980–95 are similar to post-1995 data (Fig. S2). This assumption is important, because deaths caused by building collapse depend on the corruption prevailing at the time of construction, not at the time of collapse.

Some countries are less corrupt than others with equivalent income levels (Fig. 2). We assigned these outliers an 'expectation index', between -2 and $+3$ CPI units, with negative values denoting those more corrupt than might be expected. A three-dimensional plot (Fig. 3) reveals that about 83% of all deaths from earthquakes in the past three decades have occurred in poor countries that are more corrupt than one might expect from their per capita income.

This striking correlation does not uniquely distinguish between the relative contributions of poverty and corruption, but it suggests that where corruption is extreme, its effects are manifest in the building industry. The wealthiest of nations afflicted by earthquakes can afford both to educate their populations and to purchase good-quality building materials. So it seems probable that large numbers of fatalities from earthquakes in countries below the regression line in Fig. 2 can be attributed largely to the effects of corruption. By contrast, Chile and New Zealand are less corrupt than might be expected from their per capita income, and have low earthquake fatalities. Japan, with its high per capita income and low levels of corruption, is an anomaly that we attribute to the collapse of older structures in Kobe that predate the adoption of a code of earthquake-resistant building.

STARK REMINDER

In sum, there is statistical support for widespread anecdotal evidence of a correlation between corruption and loss of life in earthquakes. Haiti and Iran are extreme examples of nations where fatalities from earthquakes are excessive and where perceived levels of corruption are above average.

"The global construction industry is the most corrupt segment of the world economy."

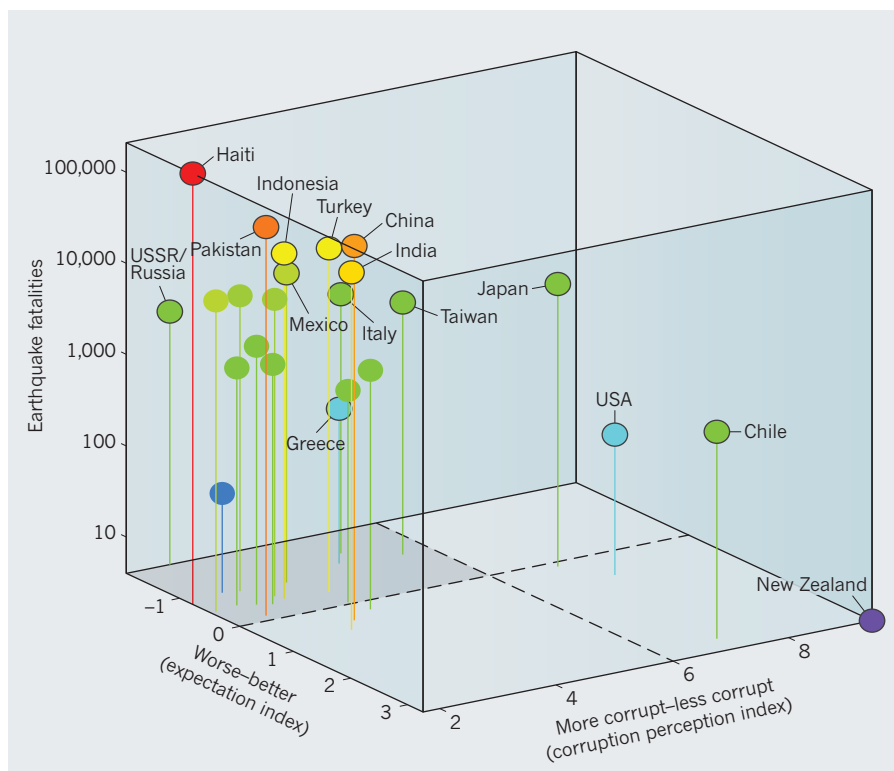


Figure 3 | Corruption's toll. Corruption versus the level of corruption that might be expected from per capita income. Of all earthquake fatalities attributable to building collapse in the past three decades, 82.6% occur in societies that are anomalously corrupt (left-hand corner of the plot).

The statistics also support last year's widely voiced opinions that the probability of earthquake-related deaths is less a function of geography and more the ability to afford earthquake-resistant construction and to enforce building codes.

Sadly, these figures have no predictive value. Moreover, even if corrupt practices were eliminated, many present-day impoverished nations will have inherited a building stock that to some degree incorporates the products of corrupt practices. The problem of what to do about these existing poorly built constructions is particularly difficult, if not economically insoluble.

But our analyses suggest that international and national funds set aside for earthquake resistance in countries where corruption is endemic are especially prone to being siphoned off. The structural integrity of a building is no stronger than the social integrity of the builder, and each nation has a responsibility to its citizens to ensure adequate inspection. In particular, nations with a history of significant earthquakes and known corruption issues should stand reminded that an unregulated construction industry is a potential killer. ■

Nicholas Ambraseys is in the Department of Civil and Environmental Engineering, Imperial College London, London SW5 2BU, UK. **Roger Bilham** is at the Cooperative Institute for Research in Environmental

Sciences and the Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309, USA.

e-mails: rogerbilham@googlemail.com, n.ambraseys@imperial.ac.uk

- Green, P. Br. J. Criminol. **45**, 528–546 (2005).
- Lewis, J. in *Global Corruption Report 2005* 23–30 (Transparency International, 2005).
- Escaleras, M., Anbarci, N. & Register, C. A. *Public Choice* **132**, 209–230 (2007).
- Bilham, R. *Bull. Earthq. Eng.* **7**, 839–887 (2009).
- Ambraseys, N. Transparency and earthquake losses *Proc. Acad. Athens* **85**, Rep. 10.06.2010 (2010).
- Burton, I. & Kates, R. W. *Nat. Resour. J.* **3**, 412–441 (1964).
- Jackson, E. L. & Burton, I. in *The Assessment and Mitigation of Earthquake Risk* 241–260 (UNESCO, 1978).
- Transparency International 2010 *Corruption Perceptions Index* (2010); available at go.nature.com/znxqt9
- World Bank, GNI per capita, Atlas method; available at <http://go.nature.com/ucv9Ue>
- Global Construction 2020: A Global Forecast for the Construction Industry over the Next Decade to 2020* (Global Construction Perspectives and Oxford Economics, 2010).
- Ambraseys, N. in *The Assessment and Mitigation of Earthquake Risk* 140–154 (UNESCO, 1978).
- Kaufmann, D. in *The Political Economy of Corruption* (ed. Jain, A. K.) Ch. 7 (Academic, 1998).
- Allen, T. I., Marano, K. D., Earle, P. S. & Wald, D. J. *Seism. Res. Lett.* **80**, 57–62 (2009).
- Marano, K. D., Wald, D. J. & Allen, T. I. *Nat. Hazards* **52**, 319–328 (2010).
- Lomnitz, C. *Bull. Seism. Soc. Am.* **60**, 1309–1313 (1970).

Supplementary Information accompanies this article online at go.nature.com/hvewfl



WELLCOME IMAGES

The lack of standardization in the collection and storage of medical specimens (pictured) can hinder subsequent research.

Bring on the biomarkers

The dismal patchwork of fragmented research on disease-associated biomarkers should be replaced by a coordinated ‘big science’ approach, argues **George Poste**.

If researchers could establish correlations between diseases and changes in biomarkers, the ability of physicians to diagnose disease and tailor treatments to individuals would be radically improved¹. However, research into biomarkers — disease-associated molecular changes in body tissues and fluids — hasn’t yet delivered on its promise. Technologies such as proteomics and DNA microarrays have contributed a voluminous literature of more than 150,000 papers documenting thousands of claimed biomarkers, but fewer than 100 have been validated for routine clinical practice. This dismal record reflects the failure of researchers to embrace a coordinated systems-based approach.

Many chronic diseases involve changes in multiple molecular pathways², but validating associations between diseases and large sets of biomarkers is hugely challenging. Most biomarker discovery is conducted in academic labs, which tend to lack the resources and multidisciplinary expertise needed to establish robust correlations between biomarkers and people’s health status or

responses to treatment. Also, getting candidate biomarkers into large-scale validation studies poses substantial logistical and regulatory challenges. Overcoming them requires the integration of diverse skills.

Changes are needed to standardize methods and obtain the large sample sizes necessary for validation trials. The traditional model of investigator-initiated research

“Too many researchers rely on whatever specimens they can obtain conveniently.”

must be replaced with the collaborative approaches typical of ‘big-science’ projects — such as The Cancer Genome Atlas initiative of the US National Institutes of Health (NIH) to catalogue the genomic changes found in cancers. Biomarker discovery should be a component of large research networks, involving industry and experts in molecular biology, genetics, analytical chemistry, computation, engineering, clinical-trial design, epidemiology, statistics, regulation and health-care economics.

Biomarkers have long been hailed as the

key to better patient care and lower medical costs³. The American Society of Clinical Oncology, for example, estimates that routinely testing people with colon cancer for mutations in the *K-RAS* oncogene would save at least US\$600 million a year⁴. It would also spare patients futile and potentially toxic treatments — for example, people with these mutations don’t respond to drugs that inhibit epidermal growth factor receptors, which cost up to \$100,000 per treatment.

APPLES AND ORANGES

A major impediment to progress in the hunt for biomarkers is the lack of standardization in how specimens are collected. Unless specimens are taken from people who are matched for as many variables as possible, all the subsequent steps in efforts to correlate biomarkers with people’s conditions and responses to treatments are compromised. At the very least, people should be matched for sex, age, weight, ethnicity, lifestyle factors such as smoking and alcohol use, and the previous treatments they’ve received. In practice, differences in how patients are

assessed, and in how those assessments are reported by different physicians in different medical centres and in different countries, mean that this rarely happens.

Too many researchers rely on whatever specimens they can obtain conveniently from local institutions. These might be fixed tissues from pathology laboratories, the usefulness of which is often limited by inadequate medical records or a lack of donor consent. Much biomarker research is carried out on cultured cell lines as surrogates for patient tissue samples. These studies can be valuable, but only if researchers demonstrate that the properties of their cell lines match those of the equivalent cell type in the human body.

The standardization problem also applies to how specimens are handled and stored, which can dramatically affect the levels of biomarkers detected. Such details are rarely documented. In a 2009 NIH survey, researchers from 80% of more than 700 responding laboratories said they struggled to obtain standardized specimens for biomarker research. Alarmingly, a similar percentage did not question how specimen quality or handling conditions might affect their results⁵.

Another pervasive problem is insufficient sampling. Most biomarker studies examine fewer than 100 specimens and lack the statistical power needed to demonstrate a robust association between multiple biomarkers and a particular condition⁶. Depending on how many biomarkers are profiled, hundreds, or even thousands, of matched control and disease samples may be needed to satisfy regulatory requirements and demonstrate that a test confers clinical or economic benefits.

BEST PRACTICE

The recent shift to testing for multiple biomarkers (multiplex profiling) has stimulated impressive innovation in high-throughput technologies for automated parallel profiling of genes, proteins, RNA molecules and metabolites. But such advances will mean little unless organizational and funding reforms persuade the research community to adopt common standards and a cross-disciplinary, systems-based approach to biomarker discovery and validation.

Several initiatives are attempting to improve the practice of tissue storage or biobanking⁷. For example, the US National Cancer Institute's Cancer Human Biobank (caHUB) has established stringent guidelines to ensure that samples from healthy individuals and cancer patients are collected, annotated, stored and analysed under standardized conditions and accompanied by appropriate donor medical information. Such initiatives offer a 'best practice' model for biobanks in general. There is also a need for the creation of international biobanks, which would require harmonized rules to allow researchers worldwide to access the resources, and

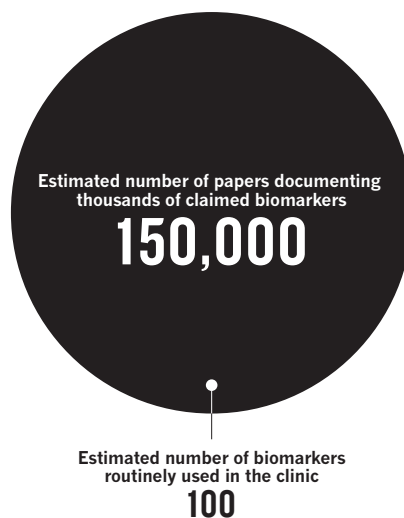
not be hindered by incompatible standards or uncoordinated regulatory barriers.

Most important, funding agencies, such as the NIH, the European Framework Programme, the Innovative Medicines Initiative (also in Europe) and the UK Wellcome Trust, should demand better returns for their investments in biomarker research. They should support only research programmes that: have access to a sufficient number of stringently characterized specimens; impose rigorous quality control in specimen acquisition, handling and storage; and possess the full spectrum of cross-disciplinary capabilities needed to translate laboratory findings to the clinic. This would encourage academic laboratories to become part of larger research networks that include clinical and industrial partners.

The large sample sizes required to validate multiplex tests are already generating massive data sets. As whole-genome sequencing becomes commonplace, the data sets will only get bigger⁸. Analysing them will require the sophisticated mathematical, statistical and computational skills of specialists in informatics. For such large-scale projects, data must also be documented, stored and analysed in standardized ways. To get candidate biomarkers into the clinic will require

A DROP IN THE OCEAN

Few of the numerous biomarkers so far discovered have made it to the clinic.



making the data formats used in academic research compatible with the formats used by industry and regulators, and with the software used to produce electronic patient health records. Industrial and clinical partners can be a powerful force in driving such standardization.

As candidate biomarkers progress to expensive validation trials, industrial partners will be essential. They can provide financial support and expertise in large-scale assay production, clinical-trial design, data analysis and regulatory compliance.

For complex multiplex tests, it is crucial that regulation experts are involved in the design of biomarker validation trials, to advise on what regulatory agencies are likely to demand. The hurdles that several biotechnology companies have encountered recently after marketing genetic tests for disease risk without first setting standards for their tests indicate the peril of not doing this⁹.

The validating process will also need to involve experts in health-care economics. In an increasingly cost-conscious health-care environment, regulatory approval will not be a guarantee of clinical adoption. Researchers must prove that their diagnostic tests will change clinical practice and reduce costs, either by improving people's health or by eliminating ineffective, expensive treatments¹⁰.

The obstacles to progress in biomarker research reside as much in the culture and organization of academic research as in deficiencies in analytical technologies. To become clinically useful, biomarker research must operate more like the large, collaborative networks mobilized for international genome-wide association studies and the multi-institution, multi-investigator big-science projects conducted in physics, climate change and complex-systems modelling.

The ability of biomarkers to improve treatment and reduce health-care costs is potentially greater than in any other area of current medical research. Thousands of papers have been written, but too few clinically useful biomarkers have been produced given a global public investment of several hundred million dollars over the past decade. It is time to restructure the field to focus instead on delivering tangible advances in health care. ■

George Poste is chief scientist at the Complex Adaptive Systems Initiative, and Regents professor of health innovation at Arizona State University, Scottsdale, Arizona 85257, USA.
e-mail: george.poste@asu.edu

1. Auffray, C., Chen, Z. & Hood, L. *Genome Med.* **1**, 2 (2009).
2. Shadt, E. E. *Nature* **461**, 218–223 (2009).
3. Davis, J. C. *et al. Nature Rev. Drug Discov.* **8**, 279–286 (2009).
4. go.nature.com/ulasjk
5. Vaught, J. *Research & Policy Initiatives in NCI's Office of Biorepositories & Biospecimen Research* (Office of Biorepositories & Biospecimen Research, 2009); available at go.nature.com/61yq9m
6. Ransohoff, D. F. & Gourlay, M. L. *J. Clin. Oncol.* **28**, 698–704 (2010).
7. Moore, H. M. *et al. Cancer Res.* **69**, 6770–6772 (2009).
8. Biesecker, L. G. *et al. Genome Res.* **19**, 1665–1674 (2009).
9. Javitt, G. *Nature* **466**, 817–818 (2010).
10. Meckley, L. M. & Neumann, P. J. *Health Policy* **94**, 91–100 (2010).

The author has declared a conflict of interest, which is available online at go.nature.com/mften7



Access to coastal waters by research vessels is being restricted by nations who exert ownership rights.

OCEANS

The blue frontier

A collection of essays highlights the pressing challenges of managing global waters, finds **Clive Schofield**.

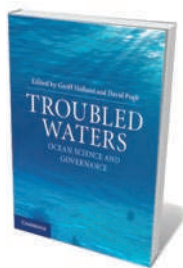
As science-fiction author Arthur C. Clarke noted, it is “inappropriate to call this planet Earth when it is quite clearly Ocean”. Crucially important to the global environment, oceans are threatened by escalating pollution, population and coastal development on land. Global climate change is contributing to sea-level rise, warming and acidification. Marine biodiversity is falling, while occurrences of cyclones and storm surges are rising. Yet the oceans also offer tremendous promise for our future resource needs.

In this context, *Troubled Waters* is timely. A diverse collection of essays by international experts in the science and policy of marine protection, its publication marks the 50th anniversary of the Intergovernmental Oceanographic Commission (IOC), a programme of the United Nations Educational, Scientific and Cultural Organization (UNESCO).

Although great progress has been made in marine science and ocean governance, the authors emphasize that advances are inadequate given the scale, complexity and urgency of the challenge. Future pressures will

exacerbate threats and opportunities, calling for even greater understanding of the seas.

The authors highlight achievements as well as problems, including shortcomings of the IOC. Longevity and success has led to the body being termed a “permanent miracle” by one contributor. But the IOC’s lack of independence from governments and the UN is a source of ambiguity and tension. Authors note the difficulty of maintaining financial commitment from participating nations, and the IOC’s lack of power to force countries to implement its recommendations. According to one account, the commission’s own successful ocean-mapping programme is “facing extermination” because of funding reductions by some member states.



Troubled Waters:
Ocean Science and
Governance
EDITED BY GEOFF
HOLLAND AND DAVID
PUGH
Cambridge University
Press: 2010. 330 pp.
£30

Oceans can be hostile neighbours, as highlighted by disasters such as the 2004 Boxing Day tsunami in the Indian Ocean and the November 1970 cyclone in East Pakistan (now Bangladesh), in which a 9-metre-high storm surge inundated the low-lying islands of the Ganges delta and caused more than 300,000 deaths. Thoughtful contributions in the book address efforts to build local marine-science capacity in developing countries in order to better detect and adapt to such threats, and describe how responses to marine hazards and early warning systems might be improved.

A theme throughout is the culture clash between scientists, who deal in probabilities, and policy-makers, who desire certainty. The book includes practical tales from scholars and practitioners who have attempted to translate ocean science into better governance. The transnational nature of marine governance is a major issue.

Several chapters refer to the oceans as a global commons that provides essential ecosystem services to support life on Earth. Yet access to large areas is being restricted. Those parts of the ocean along coasts fall under national jurisdiction as part of the territorial sea (usually extending 22 kilometres or 12 nautical miles from shore) or are defined as exclusive economic zones (EEZs). These extend 370 kilometres (200 nautical miles) from the coast, in accordance with the UN Convention on the Law of the Sea (UNCLOS), which came into force in 1994. The book acknowledges this dramatic shift in rights — the largest twentieth-century transfer of property from international to national regimes — yet more could have been said about the rising tensions between marine researchers and coastal states regarding access to the EEZs.

Coastal states increasingly impose controls in what they regard as their own waters. Scientific activity, including marine research and analogous surveys undertaken by unarmed military research vessels, has led to incidents at sea. For example, in March 2009, USS *Impeccable*, a surveillance vessel of the US Navy, was confronted by several Chinese ships off the coast of China. In 2001, India protested over the activities in its EEZ of the UK Royal Navy survey ship HMS *Scott*; the ship went on to conduct the first survey of the sea floor around the epicentre of the earthquake that caused the Boxing Day tsunami.

Similar tensions have arisen in trying to distinguish between pure and commercial marine scientific research; for exam-

NATURE.COM
How Jane Lubchenco
took on the Gulf of
Mexico oil spill:
go.nature.com/xvzzno

ple, whether marine bioprospecting is a scientific or exploitative activity. Coastal states have rights over the resources within

their EEZs. But scientific and bioprospecting activities are indistinguishable at the early sample-collecting stage. Only later, if research leads to discoveries such as drugs derived from marine genetic material, can a distinction be made. Unfortunately, this volume does not deal with these controversies in depth.

Looking to the future, *Troubled Waters* explains the high likelihood of continuing sea-level rise, increasing ocean acidification and warming oceans. This is coupled with increasing pollution and contamination

“Exploitation of the oceans is likely to accelerate as new marine opportunities are realized.”

of maritime spaces and the continuing collapse of marine biodiversity. Contributors suggest new strategies to deal with sea-level rise, including the

creation of more habitable space through reclamation, artificial islands and even mobile human habitats.

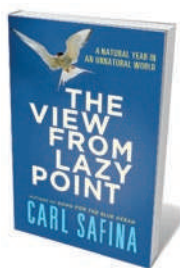
The economic worth of oceans is clearly growing. For example, the combined value of Australia's marine industries was recently found to exceed the gross value of its agricultural production. The global economy is already reliant on sea-borne trade, with more than 80% of trade by volume being carried by sea. Traditional marine resources such as fisheries and sea-bed hydrocarbons remain vital — fisheries provide the primary protein needs of hundreds of millions of people, and around 60% of global oil supplies come from offshore sources.

Exploitation of the oceans is likely to accelerate as new marine opportunities are realized, such as genetic resources, sea-floor minerals and gas hydrates. Other growth areas include ecotourism, ocean energy production and initiatives to mitigate climate change, including sequestration of carbon dioxide. Consequently, the oceans are becoming more intensively used. Activities will compete with one another in the same marine spaces, requiring enhanced governance of the oceans.

This greater economic and social interest demands increased knowledge of the oceans. The quality and quantity of ocean observations, including of relatively unexplored deep areas, should rise as technology advances. But the key to managing the seas will be turning good information into meaningful policies. In that regard, *Troubled Waters* is an excellent resource. ■

Clive Schofield is director of research at the Australian National Centre for Ocean Resources and Security (ANCORS) at the University of Wollongong, New South Wales 2522, Australia.
e-mail: clives@uow.edu.au

Books in brief



The View from Lazy Point: A Natural Year in an Unnatural World

Carl Safina HENRY HOLT 416 pp. \$32 (2011)

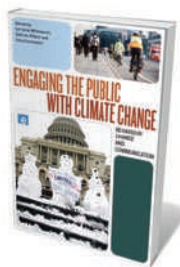
Beginning his journey in a kayak on the waters outside his Long Island beach house in New York, ocean conservationist Carl Safina witnesses the migrations of living things across the globe. Travelling from pole to pole and across the tropics during the four seasons, he brings back tales of environmental change in our seas. Although the news isn't good — reef ecosystems are being destroyed by fishing, and penguins and migrating shorebirds are starving as their food webs unravel — he remains struck by nature's beauty and versatility.



A Century of Eugenics in America: From the Indiana Experiment to the Human Genome Era (Bioethics and the Humanities)

Edited by Paul A. Lombardo INDIANA UNIVERSITY PRESS 268 pp. \$24.95 (2011)

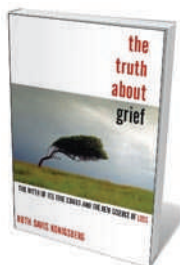
As a nation with lofty ambitions, the United States has had a mixed relationship with eugenics. The first country to prohibit procreation by criminals and 'idiots' — in the state of Indiana in 1907 — today it embraces the Human Genome Project and the possibility of genetic enhancement. Law professor Paul Lombardo examines US legislation and attitudes to human selection in the past century, and the likelihood of such pressures arising again in modern genetics.



Engaging the Public with Climate Change: Behaviour Change and Communication

Edited by Lorraine Whitmarsh, Saffron O'Neill and Irene Lorenzoni EARTHSCAN 288 pp. \$84.95 (2011)

Communicating climate science is difficult and politically fraught. A volume edited by scientists from the Tyndall Centre for Climate Change Research at the University of East Anglia, UK, examines what works, what doesn't and why. It highlights best practice from around the world using a collection of case studies from academics and practitioners, who share their advice on how to get the climate message to the public and how to promote behaviour change.



The Truth About Grief: The Myth of Its Five Stages and the New Science of Loss

Ruth Davis Konigsberg SIMON & SCHUSTER 272 pp. \$26 (2011)

Grief is often described, after psychiatrist Elisabeth Kübler-Ross, as following five stages: denial, anger, bargaining, depression and acceptance. Journalist Ruth Davis Konigsberg takes issue with this sequence and proposes a broader assessment. She points out that grieving includes positive emotions, and that we have a capacity for resilience to loss. Drawing on scientific research, she examines how people cope with grief, concluding that although psychotherapy offers support, it does not alleviate the distress experienced.



The Great White Bear: A Natural and Unnatural History of the Polar Bear

Kieran Mulvaney HOUGHTON MIFFLIN HARCOURT 272 pp. \$26 (2011)

Polar bears are as fascinating as they are striking. Born in snowdrifts, they have white fur yet black skin; they struggle to keep cool in the Arctic climate; they are massive yet pad silently on the ice; and they can wander thousands of miles in a year. Through a blend of fact, cultural history and personal experience, writer Kieran Mulvaney celebrates the paradoxical charm of polar bears, and highlights their uncertain fate as a consequence of hunting and receding sea ice.



THE ART ARCHIVE/DETROIT INST. OF ARTS/SUPERSTOCK

Henry Fuseli's *The Nightmare* (1781) sparked debate among psychologists about consciousness and dreaming.

NEUROSCIENCE

Knowing and feeling

Owen Flanagan is unconvinced by Antonio Damasio's argument that 'the self' is needed to explain consciousness.

The value of a popular science book depends on the author's ability to sort and present evidence and to be aware of its limitations. In my view, experimental neuroscientist Antonio Damasio stretches too far in his latest work, *Self Comes to Mind*, in which he considers what the self is and how it connects to consciousness.

Through his series of popular books, Damasio has become a major spokesman for a humane science of the mind that knits together reason and the emotions. His first and rightly celebrated best-seller was *Descartes' Error* (Harper, 1995). It bemoaned the continuing legacy of what in 1949 philosopher Gilbert Ryle called the "myth of the ghost in the machine", and what Damasio

judged "the abyssal separation of mind and body". His later books touched on the neuroscience behind consciousness and free will.

Self Comes to Mind is an upbeat but long-winded and desultory stroll through Damasio's musings on much of what he has thought about or worked on. With more than 90 sections on disconnected topics, each averaging a few pages, his writings come over as a neuroscientist's equivalent of haiku. The grand titles of his essays, such as 'Nature and Culture' or 'Brain and Justice', are not matched by insights. In this piecemeal collection, his thesis is hard to follow; in particular, it is not clear what he means by 'the self'.

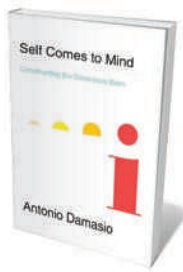
Philosophers have conveyed various views of the self over millennia: as an immaterial

soul; as some growing brain-pearl that explains feelings of self-identity; the living body; a memory connection to the past; a narrative of who I am. A vast literature has grown up around this deep psychological, philosophical and anthropological problem. Yet Damasio does not engage with it. He simply criticizes eighteenth-century philosopher David Hume's scepticism that there is a self — writing that Hume "pulverized the self" — and sides with psychologist William James's more embodied approach.

He also does not engage with the arguments and evidence supplied by cognitive scientists about the connection between consciousness and the self. Damasio offers vague signposts to his thinking: he feels that the self is a process rather than a thing; it is necessary for consciousness; and it requires a 'knower' to be generated in the brain to witness experiences in order for the mind to become conscious. Yet he does not set out scientific evidence to back up these claims.

According to most neuroscientists, consciousness refers to

➔ **NATURE.COM**
Martin Heisenberg
on free will:
go.nature.com/g4edfx



Self Comes to Mind: Constructing the Conscious Brain
ANTONIO DAMASIO
Pantheon: 2010.
384 pp. \$28.95, £25

experience. We seek to explain how consciousness arises from the brain and body and, through evolutionary biology, why it was selected for, emerged and became ubiquitous. Damasio's views on these issues are not fleshed out. As for the mechanism, he recounts the accepted picture that the brainstem, thalamus and posteromedial cortices play a big part. His 'just-so' answer on the evolutionary function of consciousness is unconvincing.

Damasio gives familiar descriptions of experiences that require an organism but do not require a 'self'. Most basic is simple sensory consciousness, as when an animal experiences pain or pleasure, hunger or thirst (which Damasio calls the proto-consciousness). Then there are more extended experiences, as when a raccoon foraging at a stream takes in the sights, smells and sensations of its surroundings (Damasio's core consciousness). And there is self-consciousness — possibly unique to humans — that includes awareness of self-referential feelings such as anxiety, or the story of one's life or character (autobiographical consciousness).

Present in all these types of consciousness, he says, is a 'self process'. This may be involved in the 'self as witness' or the 'self as protagonist', but it is not clear why it is necessary for the 'self as experiencer'. In fact, Damasio explicitly excludes dreaming — in which the sleeper has experiences but lacks self-awareness — from the set of conscious experiences. This is obviously unwarranted. We have experiences when dreaming and while under anaesthetic, even though aspects of our awareness disappear. Dreaming does not require a self process, so consciousness does not either.

Just because 'self' is in our vocabulary does not mean that it has any explanatory role in a science of the mind. Damasio says nothing convincing as to why, in addition to our fully embodied conscious beings, we ought to add the self, or self processes, to the ontological table of elements. ■

Owen Flanagan is James B. Duke Professor of Philosophy and professor of neurobiology and of psychological and brain sciences at Duke University, Durham, North Carolina 27708, USA. e-mail: off@duke.edu

NEUROSCIENCE

Mind your head

Josie Glausiusz enjoys a thought-provoking exhibition.

To step into the human brain is like being plunged into a pulsating bowl of psychedelic spaghetti. At least, that is the sensation visitors experience on entering the *Brain* exhibition at the American Museum of Natural History in New York. Bursts of light rush along a gigantic jumble of salvaged wires that hang from a ten-metre-long walkway. The installation, by artist Daniel Canogar, simulates the frenetic activity of neurons as they fire electrical impulses.

The real human brain — a plastinated example of which is on display — seems puny after such a dramatic artwork. Yet this complex organ, weighing only 1.4 kilograms and filled with 100 billion neurons, has enabled humans to colonize the planet, create great works of art, build robots and rockets, calculate and cogitate, fall in love and be conscious of our own existence. The exhibition covers all this and more.

An enchanting video captures many of the brain's functions. As a graceful young dancer performs pliés and pirouettes in an audition, a large-scale replica of the brain lights up different regions as they become active. Her cerebellum coordinates movement and balance; her limbic system triggers emotions and memories; her auditory cortex interprets music; and her motor cortex sends messages to her muscles. When she is nervous, her amygdala triggers regions of her brainstem to send signals to her heart, speeding its beat. Using her prefrontal cortex, she consciously controls her breathing to calm herself down. When she learns she has passed her test, her hippocampus allows her to save the happy memory in long-term storage.

Emotions play a big part in how our brains work, often overriding rational decision-making. There is a reason for that, says Margaret Zellner, a behavioural neuroscientist at New York's Rockefeller University and a consultant to the exhibition. "A lizard brain is a very reflexive, automatic system," she says, pointing to a stuffed green iguana on

Brain: The Inside Story

American Museum of Natural History, New York.
Until 15 August 2011.

display. By contrast, the mammal brain makes us much more flexible, thanks to our limbic system. It generates emotions that are key to our everyday survival and forms memories. If the limbic system is damaged, "our social interactions are disrupted, and we are much more impaired than if we lose the ability to speak," explains Zellner.

A panel shows some of the latest medical advances to help those who have other kinds of brain impairment. Deep brain stimulation, in which a battery-powered, implanted neurostimulator triggers electrical activity in areas of the brain that control movement, has been used to treat more than 80,000 people with Parkinson's disease. Transcranial magnetic stimulation, the fastest-growing non-drug treatment for major depression, uses a large electromagnetic coil placed against the scalp to generate painless electric currents that stimulate brain cells involved in mood control. It could some day be used to treat schizophrenia and bipolar disorder. Other devices use electrodes to restore sight by stimulating neurons in the retina, or allow people with limb paralysis or no limbs to move a cursor on a computer simply by thinking about it.

Even more awe-inspiring is another dangling tangle of copper wires, again courtesy

of Canogar, depicting the developing infant brain *in utero*. In the first five months of fetal development, an astonishing 500,000 neurons form every minute. Millions of connections continue to be made in the first few months of life, and each new skill stimulates the growth of neural links — even into adulthood and old age. Such development is the goal of exhibition curator Rob DeSalle: "I want people's brains to change as they go through this exhibition — for the better," he says. ■

Josie Glausiusz is based in New York City. e-mail: jg@planetjosie.net



Brain highlights the Kiki and Booba experiment: 98% of people agree on which name matches which shape, regardless of their native language.

AMNH/D. FINNIN

CORRESPONDENCE

Survey data are still vital to science

We have noticed a shift in the way funding agencies and journal editors are viewing observational science. It seems that the value of traditional surveys and exploration is being undermined — particularly in the field of microbial ecology. But without these surveys, scientific advances will stall.

Apart from the Census of Marine Life (www.coml.org), it is becoming increasingly difficult to obtain funding for even small-scale surveys of life. This is despite the success of large-scale observational studies such as the Global Ocean Sampling metagenomic study (D. B. Rusch *et al.* *PLoS Biol.* 5, e77; 2007), which resulted in more than 100 papers that helped to generate and test hypotheses on marine microbial communities. Later, the TARA Oceans survey (<http://oceans.taraexpeditions.org>) used a similar experimental approach, incorporating statistical design and contextual metadata.

The resulting explosion of metagenomic discovery now extends to the human digestive tract (www.metahit.eu; <http://nihroadmap.nih.gov/hmp>) and to the soil (www.terraenome.org; T. M. Vogel *et al.* *Nature Rev. Microbiol.* 7, 252; 2009). Only now, and only with metagenomics, do we have the potential to crack open the microbial black box that operates all of Earth's ecosystems (see, for example, the Earth Microbiome Project: www.earthmicrobiome.org).

Charles Darwin's skills in using meticulous survey observations as a basis for scientific theory are still just as relevant today.

Jack Gilbert Argonne National Laboratory, USA.
gilbertjack@gmail.com
Ronald O'Dor Dalhousie

University, Canada.
Timothy Vogel University of Lyons, France.

Innovation: lessons from UK policy

I disagree with parts of Bhaven Sampat's assessment of the US Bayh–Dole Act on ownership of federally funded research (*Nature* 468, 755–756; 2010).

The act was passed as an agent of economic policy, not of social policy. It aimed to reconnect academic research to the national economic infrastructure by allowing universities to obtain patents on their research and then commercialize it. It bypassed 'top-down' rules established by the government.

Sampat implies that universities grant only exclusive licences. The Bayh–Dole Act aimed to liberate universities from the rigid government prescription that permitted only non-exclusive licensing of federally funded inventions (most start-up companies require an exclusive licence, for example). In practice, the licensing mix of academic institutions is much more nuanced — 61% of US licences in fiscal year 2009 were non-exclusive (Association of University Technology Managers survey).

The United Kingdom had its 'Bayh–Dole moment' in 1988, when Prime Minister Margaret Thatcher abolished the monopoly on British academic inventions held by the National Research and Development Corporation. But outside the largest UK universities, there was very little technology-licensing activity until 1999. Then the government introduced 'third-stream' funding for knowledge transfer by universities. This essentially gave universities an economic development mandate and paid them to fulfil it.

Third-stream and university

research funding were maintained in the UK government's budget in June 2010, despite most areas of government funding receiving 10–20% cuts. A £200-million (US\$312-million) programme was also established to create a network of proof-of-concept centres, based on the German Fraunhofer Center model (www.fraunhofer.de/en).

Lessons can be learned from the UK government's acknowledgement that academic research is the long-term driver of economic growth and that sufficient support and funding is necessary to deliver on its potential.

Ashley J. Stevens Boston University, USA.
astevens@bu.edu

Evaluate impact of communication

Effective communication between scientists and non-scientists calls for special evaluation and training procedures, not simply "more and better science" (*Nature* 468, 1032–1033; 2010).

The impact of science communications on target audiences needs evaluation. And effective evaluation requires upstream planning and clear objectives if it is to inform practice (E. Jensen & B. Wagoner *Cult. Psychol.* 15, 217–228; 2009).

Unfortunately, institutions that sponsor science-communication activities are not always consistent in their evaluation requirements and rarely assess long-term impact. And when such projects are subject to independent evaluation, tenders are typically assessed either by the practitioners whose work is under scrutiny, or by staff without the relevant methodological expertise.

To enhance the quality of their engagement with the public and with policy-makers, scientists

and other science communicators should make use of theories and results from the social sciences. Without training in media literacy, audience reception or the communication and sociology of science, communicators could find themselves rehearsing long-discredited practices.

I propose that such training should be incorporated into scientists' graduate studies. This would spare governments and scientific institutions from reframing the funding and practice of science communication to protect the sciences and the public from the largely unacknowledged risks of miscommunication.

Eric Jensen University of Warwick, UK.
e.jensen@warwick.ac.uk

Hungary works on toxic sludge soil

Your revisit of last October's toxic sludge disaster in Hungary is reassuring in many respects (see go.nature.com/nhpboi). The Hungarian government is taking further steps to avert future ecological effects of the disaster.

The 800-hectare spill from an alumina factory was highly alkaline, leading to fears that arsenic and metals such as mercury and chromium could have leached into the soil and polluted the underlying water table after heavy rainfall.

Prompt governmental measures reduced the pH from 12 to 8–8.3 and rescued the River Danube's Torna creek. The government is also removing the top 15–18 centimetres of soil at the accident site. It hopes to grow a bioremediating forest of *Pteris vittata*, a fern that hyperaccumulates arsenic, to help restore soil conditions and revive the former flora and fauna.

Poonam Kaushik Bakhshi University College London, UK.
poonamkaushik@gmail.com

ANIMAL BEHAVIOUR

The price tag

Studies of animal populations often use tags to track the fate of individuals and assume that there is no adverse impact. Work on penguins shows that seemingly innocuous flipper bands affect survival and breeding success. [SEE LETTER P.203](#)

RORY P. WILSON

Biologists seeking explanations for changes in animal abundances need to be able to differentiate between animal movements and actual population change. The solution is to follow individuals and, because individuals are difficult to recognize, we need to mark them — something that is generally done with some form of animal-attached tag. Knowledge of the fate and fortunes of known animals can show how they respond to their environment, including climate change, and gives us the capacity for monitoring and prediction. What, then, is one to make of the long-term study by Saraux *et al.*¹ (page 203 of this issue), which shows that the act of tagging changes the very nature of the way animals react?

Saraux *et al.* marked king penguins (*Aptenodytes patagonicus*; Fig. 1) either with minute, subcutaneous transponders or with both transponders and the more conventional, and seemingly innocuous, flipper bands. Transponder data can be accessed only by using specialist reading equipment placed close to the tag, whereas flipper bands, the norm for decades, can be read by observers at a distance with binoculars. The bands, however, are evidently not innocuous — the authors discovered that, over 10 years, banded birds had 16% lower survival and 39% fewer chicks than the unbanded group.

Flipper bands on penguins are controversial. Different studies have reported everything from no apparent detriment to real cause for concern², although results are confounded by great variability in penguin breeding and survival, and the problem of not being able to identify a bird unless it is banded. The king penguin research¹ dealt with this by adopting a multi-year approach, and defined the effect of bands by using the presumed least detrimental tag, the transponder, as the closest to an untagged situation.

Ultimately, then, even Saraux *et al.*¹ did not deal with truly unmarked penguins, but the differences between their two bird groups indicate that flipper bands alone could affect population processes. Because many long-term population studies need extensive banding to make the data statistically robust,



Figure 1 | King penguins in the swim. The adverse effects of flipper bands could in part be due to rubbing during swimming, and increased drag.

the implications for using bands on threatened species are profound.

Flipper bands may affect penguins in several ways. For example, injury can occur to flippers and adjacent parts of the body due to rubbing during swimming³ (medium-size penguins beat their flippers about three times per second when swimming), and studies in swim channels have indicated that penguins with bands use about 24% more power to swim than birds without bands⁴. Although such studies have been dismissed as reflecting a short-term effect that birds can eliminate over time, it is hard to see how they can compensate for increased drag. A similar argument counters the theory that bands do not affect penguin mortality, because flashing bands during swimming make the birds more conspicuous to predators⁵. Habituation to newly placed animal tags is a well-documented phenomenon⁶. But we should not confuse physical impairment with psychological disturbance.

The extended timescale of Saraux and colleagues' research¹ highlights a previously undocumented complexity in the banding debate, because banded penguins suffered 30% higher mortality than unbanded birds for the first 4.5 years, with no difference between the groups after that. The explanation is not habituation, but rather a weeding out of lower-quality individuals from the population, true

to the principles of natural selection albeit exacerbated by the banding process. Here, though, the research¹ cannot be definitive because missing individuals are rarely found dead. The 'missing, presumed dead' adage may also be due to emigration; because underperforming seabirds may move colony⁷, this alone might explain some of the apparent mortality trends. But neither premature death nor induced emigration is desirable in research designed to observe rather than affect the animals being studied.

The long time-series of Saraux and co-workers' study also indicates how ocean productivity, an index of prey abundance, differentially affected banded and non-banded penguin success in breeding. The two groups fared similarly in years of high and low prey availability, but unbanded birds performed better in intermediate years. Thus, despite not being handicapped, even unbanded birds cannot breed successfully when conditions are bad; and banded birds, despite their handicap, can breed successfully when conditions are particularly good.

In normal years, however, the handicap can apparently make the difference between successful and less-successful foraging. Part of this is reflected in the lengths of foraging trips, which were greater in banded birds. Longer foraging trips cost the parents more. Such

DULL/CORBIS

trips also decrease the provisioning rate of the chicks, because adults cannot bring back more food in their stomachs to their brood just because they have been away longer.

Despite some promise in computer-enhanced automatic recognition systems based on individual markings⁸, researchers clearly must continue tagging animals for various reasons. Like non-tag marking methods, such as branding⁹ or clipping toes¹⁰, tags inevitably affect some aspect of an animal. The aspiration is that the nature of the beast is not changed by the process. However, given that selection can act on minute differences between individuals, this seems naive.

Instead, we should acknowledge that tags will impair animals. We can then strive to minimize the effects, quantifying them where possible so that we can put the resulting data into perspective. With such an informed background, proposed tagging programmes can consider whether the gain in (imperfect) knowledge

from a scientific viewpoint ethically justifies the harm inflicted. ■

Rory P. Wilson is at the Institute of Environmental Sustainability, Swansea University, Singleton Park, Swansea SA2 8PP, UK.
e-mail: r.p.wilson@swansea.ac.uk

1. Sarau, C. *et al.* *Nature* **469**, 203–206 (2011).
2. Jackson, S. & Wilson, R. P. *Funct. Ecol.* **16**, 141–148 (2002).
3. Sallaberry, M. & Valencia, J. J. *Field Ornithol.* **56**, 275–277 (1985).
4. Culik, B. M. *et al.* *Mar. Ecol. Prog. Ser.* **98**, 209–214 (1993).
5. Trivelpiece, S. G. & Trivelpiece, W. Z. *Rep. Workshop on Seabird–Researcher Interactions* Monticello, Minnesota, USA (1994).
6. Chipman, E. D. *et al.* *J. Wildl. Mgmt* **71**, 1662–1668 (2007).
7. Boulinier, T. *et al.* *Biol. Lett.* **4**, 538–540 (2008).
8. Sherley, R. B. *et al.* *Endangered Species Res.* **11**, 101–111 (2010).
9. Hastings, K. K. *et al.* *J. Wildl. Mgmt* **73**, 1040–1051 (2009).
10. May, R. M. *Nature* **431**, 403 (2004).

PARTICLE PHYSICS

Beyond Feynman's diagrams

Generations of physicists have spent much of their lives using Richard Feynman's famous diagrams to calculate how particles interact. New mathematical tools are simplifying the results and suggesting improved underlying principles.

NEIL TUROK

The world works in mathematical ways that we are occasionally privileged to discover. The laws of particle physics, for example, allow us to describe the basic constituents of the Universe, and their interactions, mathematically with astonishing precision and power. However, many important physical processes are so complicated that to perform the required calculations in traditional ways was, until recently, simply unfeasible. In near-simultaneous and complementary papers, Alday *et al.*¹ and Arkani-Hamed *et al.*² have introduced mathematical concepts that bring the calculations under control and provide insights of both immediate practical and deep theoretical importance.

The mathematical framework that we use to describe elementary particles such as electrons and photons, and their interactions, is known as quantum field theory. It was born from the synthesis of quantum mechanics with Maxwell's classical theory of electromagnetic fields and light. Unlike classical fields, quantum fields can be excited only in certain pre-specified, quantized packets of energy called particles. A photon, for example, is the elementary

particle of the quantized electromagnetic field. In the very simplest quantum field theories, the particles do not interact with each other; they merely travel singly through space at a fixed speed. But in more realistic quantum field theories, the particles collide, scatter off each other, and emit or absorb additional particles at rates that are governed by an overall parameter called the interaction coupling.

The physicist Richard Feynman developed a beautiful pictorial shorthand, called Feynman diagrams, for describing all of these processes. The diagrams show a number of initially widely separated particles moving towards each other, interacting, and flying apart again. To calculate the probability of any particular particle-interaction outcome, one draws all the contributing Feynman diagrams at each order of the interaction coupling, translates them into mathematical expressions using Feynman's rules, and adds all of the possible contributions together. This is a well-defined procedure, but at successive orders of the interaction coupling, the number of contributing diagrams grows rapidly and calculations quickly become arduous. Generations of physicists have spent large parts of their lives working out Feynman's formulae for many kinds of scattering processes, and then testing

those formulae in detailed experiments.

The work of Arkani-Hamed *et al.*² originates in a heroic, if mundane, computation undertaken in 1985 by two particle physicists at Fermilab in Batavia, Illinois. Parke and Taylor³ decided to compute all of the Feynman contributions to one of the simplest processes involving the strong nuclear force, whose elementary particle — the gluon — binds quarks together into protons and neutrons. They considered two incoming gluons colliding and producing four outgoing gluons. This is one of the most common processes: for example, in the Large Hadron Collider (LHC), located at CERN, near Geneva, Switzerland, it takes place 100,000 times per second and generates an enormous 'background' signal, which particle physicists must accurately predict and subtract as they search for signals indicating new physics.

The leading contribution to this six-gluon process involves no less than 220 Feynman diagrams, encoding tens of thousands of mathematical integrals. Yet Parke and Taylor found that they could express the final result in just three simple terms. This was the first indication that Feynman diagrams were somehow complicating the story, and that there might be a simpler and more efficient description of these scattering processes.

Further insight into this simplicity was gained by Bern, Dixon and Kosower⁴, and by Britto, Cachazo, Feng and Witten^{5,6}, who developed powerful new techniques — not involving Feynman diagrams — to infer higher-order scattering processes from lower-order ones. Their methods are not only of interest for experiments such as the LHC, but also for testing the mathematical consistency of theories such as supergravity, which is a candidate quantum field theory of gravity.

Work done over the past year has shown why these new methods are simpler than Feynman's. The formulation of quantum field theory used in Feynman's rules emphasizes locality, the principle that particle interactions occur at specific points in space-time; and unitarity, the principle that quantum-mechanical probabilities must sum to unity. However, the price of making these features explicit is that a huge amount of redundancy (technically known as gauge freedom) is introduced at intermediate steps, only to eventually cancel out in the final, physical result.

The calculations of Alday *et al.*¹ and Arkani-Hamed *et al.*² work differently. They assert relations between quantities in a new way, so that the relations are free of these redundancies and they turn out to be sufficient to define the theory. The first big surprise is that such relations exist, and the second is that they are expressed in quantities that are explicitly non-local — that is, quantities that are spread out over space and time.

Both sets of authors perform calculations within a particularly simple family of four-dimensional quantum field theories,

with interactions, known as $N = 4$ supersymmetric theories. These theories are not realistic descriptions of real-world particle physics, but they do have elementary particles such as gluons and quarks (and even Higgs bosons), and they provide a valuable testing ground for new calculational techniques.

Arkani-Hamed and colleagues² exploit a combination of twistor theory — a non-local description of space-time developed by Roger Penrose in the 1970s — and algebraic geometry to obtain a complete description of the scattering of all the elementary particles in these theories, in ascending powers of the interaction coupling. In doing so, the authors provide an excellent characterization of the scattering process when the interaction coupling is small. By contrast, Alday and colleagues¹ derive relations between non-local quantities known as Wilson loops, named after their inventor, the Nobel prizewinner Kenneth G. Wilson. The loops represent the flux of the strong nuclear-force fields through various geometrical areas. Using the powerful mathematical machinery of quantum integrability, Alday *et al.* are able to determine the behaviour of these fluxes in the limit at which the interaction coupling is large. The two sets of authors have therefore described the theory in its two opposite extreme limits — small and large coupling — and the hunt is now on for a complete description, one that is valid for any value of the interaction coupling.

Quantum field theory is the most powerful mathematical formalism known to physics, successfully predicting, for example, the magnetic moment of the electron to one part in a trillion. The recent discovery of mathematical structures that are now seen to control quantum field theory is likely to be of enormous significance, allowing us not only to calculate complex physical processes relevant to real experiments, but also to tackle fundamental questions such as the quantum structure of space-time itself. The fact that the new formulations of the theory^{1,2} jettison much of the traditional language of quantum field theory, and yet are both simpler and more effective, suggests that an improved set of founding principles may also be at hand. ■

Neil Turok is at the Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada.
e-mail: nturok@perimeterinstitute.ca

1. Alday, L. F., Gaiotto, D., Maldacena, J., Sever, A. & Vieira, P. Preprint at <http://arxiv.org/abs/1006.2788> (2010).
2. Arkani-Hamed, N., Bourjaily, J. L., Cachazo, F., Caron-Huot, S. & Trnka, J. Preprint at <http://arxiv.org/abs/1008.2958> (2010).
3. Parke, S. J. & Taylor, T. R. *Phys. Rev. Lett.* **56**, 2459–2460 (1986).
4. Bern, Z., Dixon, L. J. & Kosower, D. A. *Ann. Phys.* **322**, 1587–1634 (2007).
5. Britto, R., Cachazo, F. & Feng, B. *Nucl. Phys. B* **715**, 499–522 (2005).
6. Britto, R., Cachazo, F., Feng, B. & Witten, E. *Phys. Rev. Lett.* **94**, 181602 (2005).

STRUCTURAL BIOLOGY

Finding the wet spots

The functions of proteins are critically coupled to their interplay with water, but determining the dynamics of most water molecules at protein surfaces hasn't been possible. A new spectroscopic method promises to change that.

VINCENT J. HILSER

Proteins in cells are responsible for the vast majority of biological functions. Because life evolved in water, protein molecules are uniquely adapted to use their aqueous environments to facilitate their functions¹. Yet remarkably little is known about the interactions between solvent water and protein molecules, or how those interactions affect (or are affected by) the conformational changes at the heart of protein function. In *Nature Structural and Molecular Biology*, Nucci *et al.*² now report that nuclear magnetic resonance (NMR) spectroscopy of proteins encapsulated in reverse micelles³ — cell-like compartments in which nanometre-scale pools of water are surrounded by a membrane — can provide a comprehensive picture of how water molecules bind to proteins. This picture not only challenges current dogma about protein hydration, but also promises to illuminate key aspects of the relationship between protein and water dynamics, and of how proteins use water to perform their functions.

Early studies^{4,5} of protein–water interactions — the exchange of water molecules between a protein's surface and the surrounding bulk water — were performed in bulk solution using NMR. But because of ambiguities resulting from the timescale of the exchange process, as well as the inability to distinguish between that process and another in which labile hydrogens in the protein exchange with those in water⁶, direct experimental analysis of protein–water dynamics (hydration dynamics) was restricted to only the most long-lived of interactions. Attempts to rectify this have relied mostly on X-ray crystal structures of proteins to identify the locations of resolvable water molecules in the structure, which, in spite of well-documented reservations⁷, have generally been presumed to represent the 'hydration shell' of water molecules around the protein⁸ (Fig. 1a). Nucci and colleagues' new NMR approach² overcomes the previous experimental limitations, thus providing a comprehensive picture of the whole hydration shell around a test protein, ubiquitin.

The reverse-micelle technology used by Nucci *et al.*² was previously developed³ to overcome the protein-size limitation inherent to NMR studies — large proteins can't be studied by NMR because they tumble too slowly in solution. Encapsulation of large proteins in

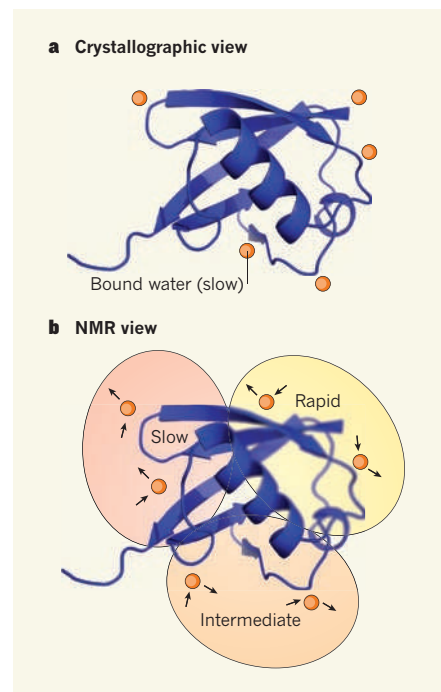


Figure 1 | Crystallographic versus NMR views of protein hydration. **a**, Certain sites at which water molecules associate with protein surfaces can be identified in X-ray crystal structures, as shown schematically here for ubiquitin. The sites are assumed to be those where water molecules reside for longest (that is, where the dynamics of water movement are slowest). **b**, This cartoon depicts the NMR view of hydration obtained by Nucci *et al.*², wherein a complete picture of the locations and dynamics of water molecules bound to ubiquitin was ascertained. They observed that water molecules cluster into regions corresponding to slow, intermediate and rapid average dynamics. Little correlation was found between the crystallographic and NMR views of hydration dynamics. Arrows indicate that the rates of water-molecule exchange between the protein's surface and the solvent are directly measured by NMR. By contrast, the X-ray picture is static, and exchange rates must be inferred or calculated. Images were created using PyMOL¹⁵.

reverse micelles dissolved in a low-viscosity fluid, however, allows them to tumble at rates similar to those of much smaller proteins. What's more, such encapsulation dramatically slows both the hydration dynamics and the hydrogen-exchange kinetics of proteins compared with the same quantities in bulk solvent. This is the cornerstone of Nucci and colleagues' advance², because it enables

the dynamics of water molecules bound to proteins to be unambiguously resolved for the first time.

The authors' key finding is that water molecules that have similar hydration dynamics form clusters across the protein surface (Fig. 1b). That is, the residence time of water on the surface is often similar at proximal regions of the protein. To understand the importance of this result, one should note that liquid water makes and breaks hydrogen bonds with other water molecules in a dynamic network⁷. The rapid molecular motions of liquid water result from cooperative rearrangements of this network — the water molecules can rearrange without going through transition states wherein one or several hydrogen bonds must break^{7,9}.

The presence of a protein can alter this network, however, changing not just the probabilities of the different structural rearrangements of solvent, but also the ease with which water molecules can interconvert between them (by altering the transition states of low-energy water arrangements). In principle, geometrical features on the protein surface can affect the residence times of individual water molecules^{10,11}: if one or many hydrogen bonds must be broken to facilitate the movement of a molecule constrained by a particular feature, the probability of such an event will be decreased, and the water molecule will tend to stay where it is. One might therefore expect the residence times of individual water molecules at the protein surface to be determined by the precise structural features of the neighbourhood of each molecule, and to be independent of the residence times at other sites.

But this isn't what Nucci and colleagues² observed for ubiquitin². Remarkably, they found that the protein confers regional effects on the solvent, so that rates of water-rearrangement processes vary by a factor of more than 10^{10} from one region of the protein's surface to another. In other words, the geometric constraints imposed by ubiquitin on the surrounding water molecules effectively create (to a first approximation) independent solvent networks around each region.

The water molecules in these regional networks seem to act cooperatively, but in a way that is blind to the behaviour of bulk solvent or of the other regions. Networks that require many hydrogen bonds to be broken in order to rearrange solvent molecules will reside longer at the surface than those that don't. Qualitatively, this observation is not altogether unexpected, as it is clear that several water molecules must be involved in any water rearrangement. It is the level at which these differences are manifested, however, that has broad implications for understanding protein function and evolution. Proteins function by recognizing and binding to ligand molecules such as cofactors, binding partners or enzymatic substrates¹². Insight into the structural and energetic factors that make a

good binding site is central to an understanding of protein function, and to the development of *de novo* protein design¹³. The fact that hydration dynamics are regionally segregated across ubiquitin suggests that some protein surfaces are more conducive to rapid solvent rearrangement than others.

Are surface-hydration properties an evolutionarily selectable trait for proteins? Certainly it is easy to imagine why enzymes, which are optimized to accelerate catalytic processes, would benefit from the evolutionary selection of rapid hydration dynamics. But are regional patterns of hydration dynamics hard-wired into the folds of particular proteins, or are they tunable through mutation? Whatever the answer, Nucci and colleagues' study² forces us to think again about the ways proteins can manipulate not just their own structure and dynamics, but also the dynamics of their solvent environment¹⁴.

It remains to be seen whether the regional patterns of hydration dynamics observed in ubiquitin also occur in other proteins. Currently, the main obstacle to applying Nucci and colleagues' analytical method² to other proteins is that preparing protein samples in reverse micelles is a non-trivial task. Nevertheless, the potential of this method to fundamentally alter our view of how proteins interact with their

solvent environment — and so also of how they function — makes the effort worthwhile. ■

Vincent J. Hilser is in the Departments of Biology and of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA. e-mail: hilser@jhu.edu

1. Eisenberg, D. & Kauzmann, W. *The Structure and Properties of Water* (Clarendon, 1969).
2. Nucci, N. V., Pometum, M. S. & Wand, A. J. *Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1955 (2011).
3. Wand, A. J., Ehrhardt, M. R. & Flynn, P. F. *Proc. Natl Acad. Sci. USA* **95**, 15299–15302 (1998).
4. Otting, G., Liepinsh, E. & Wuthrich, K. *Science* **254**, 974–980 (1991).
5. Denisov, V. P., Peters, J., Horlein, H. D. & Halle, B. *Nature Struct. Biol.* **3**, 505–509 (1996).
6. Modig, K., Liepinsh, E., Otting, G. & Halle, B. J. *Am. Chem. Soc.* **126**, 102–114 (2004).
7. Halle, B. *Phil. Trans. R. Soc. Lond. B* **359**, 1207–1224 (2004).
8. Nakasako, M. *Phil. Trans. R. Soc. Lond. B* **359**, 1191–1206 (2004).
9. Geiger, A., Kleene, M., Paschek, D. & Rehtanz, A. *J. Mol. Liq.* **106**, 131–146 (2003).
10. Chandler, D. *Nature* **437**, 640–647 (2005).
11. Sharp, K. A., Nicholls, A., Fine, R. F. & Honig, B. *Science* **252**, 106–109 (1991).
12. Creighton, T. *Proteins: Structures and Molecular Properties* (Freeman, 1993).
13. Lippow, S. & Tidor, B. *Curr. Opin. Biotechnol.* **18**, 305–311 (2007).
14. Fraunfelder, H. et al. *Proc. Natl Acad. Sci. USA* **106**, 5129–5134 (2009).
15. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2009).

CONDENSED-MATTER PHYSICS

The conducting face of an insulator

Stacking two oxide insulators together is known to yield a conducting system at the interface between the oxides. But the discovery that simply cleaving such an insulator yields the same outcome is unexpected. SEE LETTER P.189

ELBIO DAGOTTO

On page 189 of this issue, Santander-Syro *et al.*¹ report the discovery of a conducting two-dimensional electron system on the surface of an insulator, strontium–titanium oxide (SrTiO₃; or STO for short). The finding is unexpected, because STO has been much studied in the past and was believed to be fully understood; and it may have implications for all areas of research that routinely use STO, such as the rapidly growing fields of oxide superlattices and oxide electronics^{2,3}.

With its frequent discovery of new materials, condensed-matter physics is a dynamic field of research, full of surprises that regularly challenge our understanding of how electrons and atoms behave in solids. Well-known examples are high-temperature superconductors,

and transition-metal oxides (TMO) that have large magnetoresistance⁴ — a material's ability to change its electrical resistance when placed in a magnetic field. In TMO, and in bulk materials in general, the total energy of the system is minimized as a result of the atomic ions and electrons adopting a specific crystal arrangement, for example the perovskite structure, and by modifications to low-temperature properties such as their magnetic states. But it is hard to predict how a certain property will change without first preparing a sample of the compound and investigating it. And if a particular property is needed for a specific application, it is difficult to anticipate which chemical composition will produce the desired outcome. Thus, the crystal structure of bulk TMO, and its associated properties, cannot be easily controlled.

For this reason, the artificial preparation

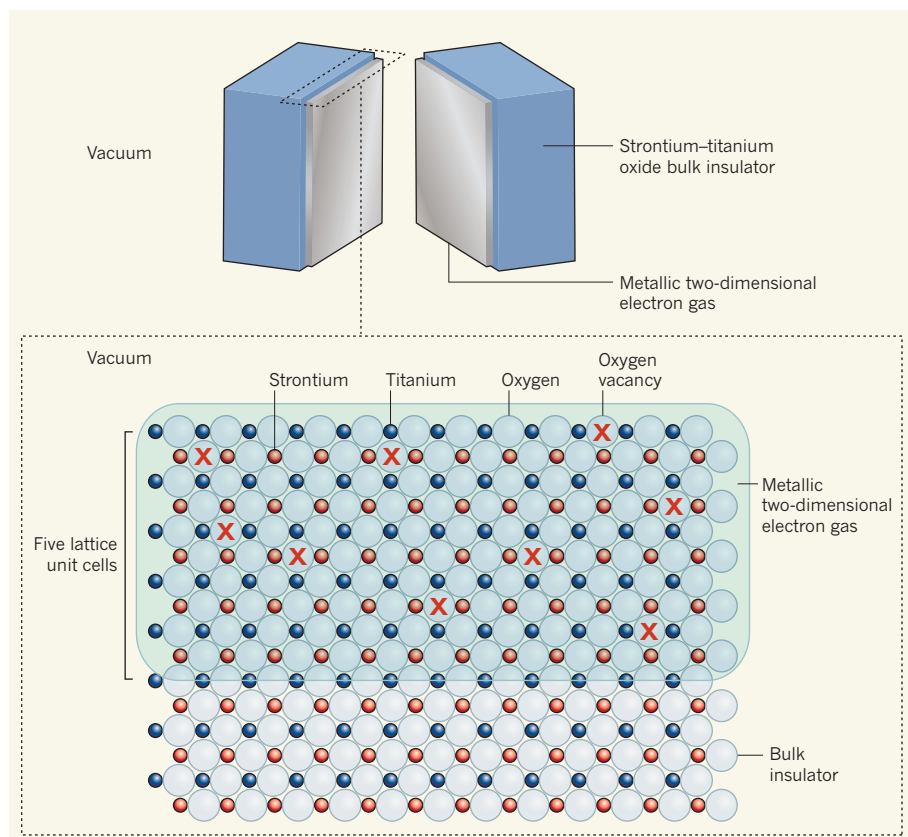


Figure 1 | Cleave the oxide. Santander-Syro *et al.*¹ create a metallic, two-dimensional electron gas on the surface of a strontium–titanium oxide (SrTiO₃) bulk insulator simply by cleaving the material in an ultrahigh vacuum. The conductivity of the electron system, confined within a depth of merely five lattice unit cells below the surface, arises from the oxygen vacancies (here arbitrarily distributed) that are created by the fracture process.

of new TMO is receiving considerable attention. These days, it is possible to prepare, with atomic precision, materials made from a combination of other materials by growing layers of a particular oxide over layers of another oxide, and repeating the process in a variety of arrangements to create superlattices. The range of combinations that can be produced is enormous, with each one leading to a novel material. The stacking of two insulating TMO, one of them being STO, has already led to the discovery^{5–7} of metallic (conducting) and even superconducting interfacial states.

At first glance, the existence of metallic states at the interface between two insulators may seem puzzling. But according to some researchers, a mechanism known as a polar catastrophe can take place that explains those metallic states. For instance, when lanthanum–aluminium oxide (LaAlO₃; LAO), which is made up of alternately charged layers of (LaO₂)[–] and (AlO)⁺, is stacked on top of STO, which is composed of charge-neutral layers of TiO₂ and SrO, polar catastrophe arises⁸ from the (LaO₂)[–]–(AlO)⁺ electric dipoles in the LAO structure. The dipoles give rise to an electric potential difference across the LAO structure that increases with the material's thickness. When the LAO structures are sufficiently

thick, a sudden electronic reconfiguration — the polar catastrophe — occurs to compensate for the large potential difference generated, conferring conductivity on the LAO/STO interface.

But what about Santander-Syro and colleagues' observation¹ of a conducting state on the surface of STO, the thin portion of the material in contact with vacuum at the top of the crystal? How can a conducting system be created on the surface of an insulating bulk compound? STO is made of electrically neutral — not charged — TiO₂ and SrO layers, preventing polar catastrophe from taking place.

In their study, Santander-Syro *et al.* investigate the properties of the surfaces of a number of STO crystals. To obtain clean surfaces, they fractured each of the crystals in an ultrahigh vacuum. The remarkable result is that the properties of the surfaces of all samples examined are those of a metallic, two-dimensional electron gas — one that is estimated to be confined within a depth of merely five lattice unit cells below the surfaces (Fig. 1). What's more, despite the fact that the samples' low electronic densities in the bulk spanned a large range of values, the authors observed the same high electronic density at the metallic surface of all samples.

These results are surprising. Santander-Syro and colleagues argue that the fracture process creates a large number of oxygen vacancies on the material's broken surface (Fig. 1). Because oxygen in the crystal adopts an ionic state (O^{2–}), the removal of each oxygen atom effectively 'returns' two electrons back to the lattice — in the sense that the ion's two excess electrons remain in the material close to the vacancies, and thus close to the surface, and are therefore available for conduction. Although the samples had quite different numbers of bulk charge carriers (introduced via impurity doping), all samples were fractured in a similar manner — implying that the samples' resulting density of oxygen vacancies might be similar and account for the observed, universal behaviour.

With a view to applications in oxide electronics^{2,3}, the properties of the conducting states present at the surface of STO and at interfaces such as LAO/STO should now be compared to clarify their origin. In addition, and as was done for LAO/STO, field-effect transistors — the building blocks of electronic chips — based on the new conducting state at the surface of STO should be considered and implemented.

Perhaps the most important implication of Santander-Syro and colleagues' results is that a new conducting, two-dimensional electron system on a familiar and popular substrate is now available by simply cleaving it — or, more generally, by producing oxygen vacancies right at its surface. This simple approach is an attractive alternative to the techniques of growing superlattices, which are complex and expensive. The procedure brings to mind analogies to the way in which the field of graphene electronics started — with the discovery that graphene (a one-atom-thick, honeycomb-like lattice of carbon) could be produced simply by exfoliating graphite using a common adhesive tape. An interesting challenge for the future of oxide electronics is to understand and control the vacuum-fractured oxide conducting system described here, and to learn how to preserve it. ■

Elbio Dagotto is in the Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee 37996, USA, and in the Materials Science and Technology Division, Oak Ridge National Laboratory, Tennessee. e-mail: edagotto@utk.edu

1. Santander-Syro, A. F. *et al.* *Nature* **469**, 189–193 (2011).
2. Mannhart, J. & Schlom, D. G. *Science* **327**, 1607–1611 (2010).
3. Cen, C., Thiel, S., Mannhart, J. & Levy, J. *Science* **323**, 1026–1030 (2009).
4. Dagotto, E. & Tokura, Y. *MRS Bull.* **33**, 1037–1045 (2008).
5. Ohtomo, A., Muller, D. A., Grazul, J. L. & Hwang, H. Y. *Nature* **419**, 378–380 (2002).
6. Ohtomo, A. & Hwang, H. Y. *Nature* **427**, 423–426 (2004).
7. Reyren, N. *et al.* *Science* **317**, 1196–1199 (2007).
8. Hwang, H. Y. *Science* **313**, 1895–1896 (2006).

CONSERVATION

The trouble with bumblebees

A survey of bumblebees in North America provides unequivocal evidence that four previously common and abundant species have undergone recent and widespread population collapse. Various explanations remain possible.

MARK J. F. BROWN

“To make a prairie it takes a clover and one bee, One clover, and a bee. And revery. The revery alone will do, If bees are few.”¹ Bumblebees, the main pollinators of red clover and much else besides, comprise a genus of some 250 species that are central both to natural ecosystem function and to a multimillion-pound commercial pollination industry involving the production and global traffic of bumblebee colonies². Unfortunately, bumblebees may indeed be becoming ‘few’. Recently there have been disturbing reports of rapid declines in North American species, with speculation that the declines are being driven by introduced parasites³.

Writing in *Proceedings of the National Academy of Sciences*⁴, Cameron *et al.* now provide evidence of geographically widespread and temporally rapid reductions in the distribution and abundance of four bumblebee species across the contiguous United States. The species, which constitute about 10% of the regional bumblebee species, had both lower genetic variability and significantly higher prevalence (the percentage of the population infected) of the microsporidian parasite *Nosema bombi* — a type of fungus that has been implicated as a potential cause of bumblebee decline.

Bumblebee decline is not a new story. Since the problem was first recognized in Britain in the 1950s, numerous studies have documented long-term deteriorations in species ranges, with habitat loss being the generally accepted cause². What makes reports from North America both worrying and intriguing is the apparent speed of the deterioration.

To test the rapidity and extent of the declines, Cameron *et al.*⁴ generated a database of historical collections from across the United States and used it, together with niche modelling software, to estimate species distribution and patterns of relative abundance since 1900 for eight target species, half of which had previously been suggested to be in decline. Current distribution and relative abundance were determined by sampling at 382 sites across 40 states between 2007 and 2009. The range of four species — *Bombus affinis*, *B. occidentalis*, *B. pensylvanicus* (Fig. 1) and *B. terricola* — had decreased by 23–87%, with the other four



Figure 1 | *Bombus pensylvanicus* — in decline.

species being, as expected, present in most of their historical range. The affected species also exhibited rapid declines — over the past 20–30 years — in relative abundance.

As the authors acknowledge, local species extinction is notoriously hard to prove, and the over-representation of rare species in historical collections makes determining relative abundance problematic. Such difficulties mean that the exact extent and rate of each decline reported in this study is open to question. But the basic result — that previously common and abundant bumblebee species have undergone recent, widespread population collapse — seems undeniable. So, what is driving these declines?

Emergent diseases are increasingly being recognized as threats to both humans and native species⁵. Pathogen transmission from commercially bred colonies to natural populations has already been seen in two other parasites of bumblebees⁶. Rapid falls in numbers of North American bumblebees, redolent of an epidemic, were contemporaneous with the collapse in commercial breeding of *B. occidentalis* in North America, which was blamed on *N. bombi*. These observations led to the hypothesis that *N. bombi*, introduced from Europe by means of commercial pollinators to native bumblebees, was driving declines³. Cameron and colleagues show that *N. bombi* indeed has significantly higher prevalence in the rapidly failing species. Furthermore, DNA sequencing demonstrated that these North

American parasites were genetically identical to European isolates.

Is this the smoking gun behind North American bumblebee declines? To answer this question, we need to know what high parasite prevalence means.

There are two obvious interpretations. First, high parasite prevalence may represent the moving edge of a wave of infections, indicating that these bumblebee populations are on the verge of extinction. Similar patterns in the fungus responsible for global amphibian declines, albeit in the intensity (number of parasite cells per individual) rather than prevalence of infection, precede local extinction of amphibians⁷. Second, high prevalence may simply indicate that the declining species naturally support high populations of the parasite.

Is there any evidence to distinguish between these interpretations? Alaskan populations of *B. occidentalis*, which remain abundant, also have a high prevalence of infection by *N. bombi* (J. P. Strange, personal communication), but this could support either explanation. Intriguingly, a parasitic mite of bumblebees has higher prevalence in species of the subgenus *Bombus* in Canada⁸, the same subgenus to which three of the four declining species belong. This supports the idea that the species are natural reservoirs for the microsporidian. By contrast, the genetic identity between North American parasites and European isolates is evidence for the emergent-disease interpretation (although large-scale genetic studies are needed to confirm this).

A final ambiguity in understanding the parasite data is that Cameron *et al.*⁴ also found lower genetic diversity in the declining species. Low diversity is predicted to increase susceptibility to parasites, and a recent study demonstrated a correlation between inbreeding and the prevalence of a trypanosome parasite in bumblebees⁹. Untangling the causal direction of correlations between patterns of decline, parasite prevalence and loss of genetic diversity in North American bumblebees will take considerable work.

This study⁴ is the first step towards understanding declines in North American bumblebees, but it also has broader implications. The methodology can be used to track declines in other species for which long-term recording schemes do not exist, a situation that applies to most of the planet's biodiversity. More specifically, global transport of commercial bumblebee colonies sets the stage for pathogen transmission to native species. If *N. bombi* has driven North American declines — and ascertaining this will require further investigation — it and other parasites may have the potential to drive native bumblebee decline across the world.

At a meeting held by the International Union for Conservation of Nature in St Louis, Missouri, in November 2010, commercial producers, non-governmental organizations, federal agencies and scientists discussed measures for conserving native bumblebee species

THOMAS WILSON

while maintaining them as essential commercial pollinators. With due respect to Emily Dickinson¹, 'revery' will not be enough if we want to see prairies, and other important terrestrial ecosystems, thriving in the future. ■

Mark J. F. Brown is in the School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK.
e-mail: mark.brown@rhul.ac.uk

1. Dickinson, E. *The Complete Poems of Emily Dickinson* (Little, Brown, 1924).
2. Williams, P. H. & Osborne, J. L. *Apidologie* **40**, 367–387 (2009).
3. Thorp, R. W. & Shepherd, M. D. in *Red List of Pollinator Insects of North America* CD-ROM Version 1 (eds Shepherd, M. D., Vaughan, D. M. & Black, S. H.) (Xerces Soc. Invertebrate Conserv., Portland, Oregon, 2005); www.xerces.org/Pollinator_Red_List/Bees/Bombus_Bombus.pdf
4. Cameron, S. A. et al. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1014743108 (2011).
5. Keesing, F. et al. *Nature* **468**, 647–654 (2010).
6. Stout, J. C. & Morales, C. L. *Apidologie* **40**, 388–409 (2009).
7. Vredenburg, V. T., Knapp, R. A., Tunstall, T. S. & Briggs, C. J. *Proc. Natl Acad. Sci. USA* **107**, 9689–9694 (2010).
8. Otterstatter, M. C. & Whidden, T. L. *Apidologie* **35**, 351–357 (2004).
9. Whitehorn, P. R., Tinsley, M. C., Brown, M. J. F., Darvill, B. & Goulson, D. *Proc. R. Soc. B* doi:10.1098/rspb.2010.1550 (2011).

EARTH SCIENCE

A back-arc in time

The Eastern Lau spreading centre in the Pacific Ocean is the subject of especial interest. The influence of the neighbouring subduction zone is considerable, but evidently has unexpected limits. SEE LETTER P.198

PETER MICHAEL

In plate-tectonic theory, ocean crust and the associated lithosphere are recycled back into Earth's mantle at the destructive plate boundaries called subduction zones. Several subduction zones also have submarine spreading centres that occur on the overriding plate lying behind the arc of surface volcanoes to be found above the site of subduction. These 'back-arc' spreading centres are the most rapidly changing plate-tectonic boundaries on the planet. New ocean crust is constructed by sea-floor spreading at back-arc spreading centres, just as occurs at mid-ocean ridges. But this spreading propels the back-arc spreading centre over the chemically diverse mantle of the subduction zone, and eventually away from the supply of subducted material that feeds the spreading.

On page 198 of this issue¹, Dunn and Martinez describe a study of crustal thickness and structure at the Eastern Lau spreading centre (ELSC) in Tonga. Their work shows that back-arc spreading centres change even more rapidly than previously thought, suggesting that they are more active in capturing the subducted input from the mantle, and then rapidly releasing most of it when the spreading centre reaches a critical distance from the arc. Figure 3a of the paper (page 201) is a map of the region: the Tonga trench is the subduction zone's intersection with the surface; triangles on the Tonga ridge show the associated volcanic arc; and the location of the ELSC is marked.

The key ingredient in subduction zones is the mineralogically bound water that is carried into the mantle in the downgoing, subducted slab and then released into the overlying mantle wedge as the cold slab is heated. It promotes greater extents of mantle melting and the production of magmas that are progressively richer in silicon dioxide (SiO₂) and water.

This results in crust that is thicker, seismically slower and more porous. It is these changes that are observed, in both crustal properties² and rock composition³, southwards along the 'zero-age' axis of the ELSC, as the distance between the Tonga volcanic arc and the ELSC diminishes and the input of subducted materials to the back-arc increases.

By examining the crustal structure across the axis as well as along it, Dunn and Martinez¹ are peering back to a time when the back-arc basin was narrower. The volcanic morphology at the surface and the seismic velocities of the underlying few kilometres of crust show that, over a short period of time, as the back-arc spreading centre pushed itself away from the volcanic arc by sea-floor spreading, the volcanic crust abruptly became smoother, thinner, denser and probably less porous. In other words, it became less influenced by subducted water. These relatively shallow observations of the crust reflect what is happening in the deeper mantle wedge.

The abruptness of the changes is the crucial factor here, as it suggests that the spreading centre is doing more than merely sampling whatever mantle it passes over. It remains to be seen how the concept of active capturing of subduction-influenced mantle and its rapid release at a critical distance will influence the increasingly sophisticated models that have been proposed for the formation of magmas behind the volcanic arc^{4,5}. The conceptual cartoons that arise from these models of magma genesis are not yet sufficiently detailed. At the same time, geophysical imaging of the mantle wedge in other arcs⁶, and geodynamic models of the mantle wedge and slab that include dehydration and rheological changes⁷, are leading to more realistic models of subduction-zone processes and hint at a region in the mantle where conditions change rapidly over short distances. The time constraints from the

ELSC provided by Dunn and Martinez¹ could improve the construction of these geodynamic models.

The authors' investigation¹ is part of the Ridge 2000 programme sponsored by the US National Science Foundation⁸. This is an interdisciplinary initiative to study Earth's oceanic spreading ridge system as an integrated whole, from its inception in the mantle to its manifestations in the biosphere and the water column. Intensive studies at three integrated study sites (including the ELSC) seek to establish links between different parts of these complex systems 'from mantle to microbe'. The ELSC was chosen as a site because of the gradational nature of the effects of subduction (especially of water) along its axis. Hydrous magma degassing and crustal composition control the composition of hydrothermal fluids⁹, and therefore also strongly influence the microfauna and macrofauna at hydrothermal vents along the spreading centre. In their work, Dunn and Martinez exploited the expected link between crustal properties and mantle-source composition.

Their research was made possible by the continually increasing investment in, and improvement of, ocean-bottom seismometers. Seismic-imaging studies use arrays of seismometers as receivers to provide a three-dimensional view of travelling seismic waves, whether fast or slow. The deployment of seismometers on the sea floor is not new. But this study¹ involved the largest, densest array of ocean-bottom seismometers deployed over an oceanic spreading centre anywhere on Earth, and permitted large-scale questions to be addressed at the ELSC.

Dunn and Martinez¹ used ship-borne airguns as the seismic-wave source to produce many relatively low-energy bursts that allowed the shallow crustal structure to be examined. Other seismic-imaging studies are under way with much longer deployments of the same seismometers, and using earthquake energy as a high-energy seismic-wave source to image the deeper mantle wedge and subducted slab. These investigations cover the same geographical area as the current shallow study, and may provide additional tests of the hypothesis of a critical distance in which volcanic-arc material is captured.

Further tests of the Dunn and Martinez hypothesis will be forthcoming. If, as required by the hypothesis, there is an excellent

correlation between crustal properties and magma composition, there should be predictable, stepwise decreases in subduction-related elements such as barium across the axis as the rocks get younger. There will undoubtedly be further expeditions to sample rocks across the axis of the ELSC to test this hypothesis. Along-axis changes in 'zero-age' rock chemistry are stepwise, but are not perfectly related to the distance between the ELSC and the volcanic arc³, suggesting a more complicated relationship.

The rapid changes in the effects of subduction proposed by Dunn and Martinez should not be limited to the ELSC, but should be common in other subduction systems in which the back-arc spreading centre has migrated farther away from the volcanic arc. For example, detailed bathymetric and seismic studies of another back-arc spreading centre in the Pacific, the Mariana system, should show similar, sharply bounded domains. ■

Peter Michael is in the Department of Geosciences, The University of Tulsa, Tulsa, Oklahoma 74104, USA.
e-mail: pjm@utulsa.edu

1. Dunn, R. A. & Martinez, F. *Nature* **469**, 198–202 (2011).
2. Jacobs, A. M., Harding, A. J. & Kent, G. M. *Earth Planet. Sci. Lett.* **259**, 239–255 (2007).
3. Escrig, S., Bézous, A., Goldstein, S. L., Langmuir, C. H. & Michael, P. J. *Geochim. Geophys. Geosyst.* **10**, doi:10.1029/2008GC002281 (2009).
4. Langmuir, C. H., Bézous, A., Escrig, S. & Parman, S. W. in *Back-Arc Spreading Systems: Geological, Biological, Chemical, and Physical Interactions* (eds Christie, D. M., Fisher, C. R., Lee, S.-M. & Givens, S.) *Geophys. Monogr. AGU* **166**, 87–146 (2006).
5. Kelley, K. A. *et al.* *J. Geophys. Res.* **111**, doi:10.1029/2005JB003732 (2006).
6. Rychert, C. A. *et al.* *Geochim. Geophys. Geosyst.* **9**, doi:10.1029/2008GC002040 (2008).
7. van Keken, P. E. *et al.* *Phys. Earth Planet. Inter.* **171**, 187–197 (2008).
8. www.ridge2000.org
9. Mottl, M. A. *et al.* *Geochim. Cosmochim. Acta* (in the press).

SYNTHETIC BIOLOGY

Division of logic labour

Cellular compartmentalization is an effective way to build gene circuits capable of complex logic operations, in which binary inputs are converted into binary outputs according to user-defined rules. SEE LETTERS P.207 & P.212

BOCHONG LI & LINGCHONG YOU

Computation forms the foundation of electronic devices that pervade our daily lives. Many of these devices run on digital circuits assembled from logic gates. All logic gates use an unambiguous rule to convert inputs of 0 or 1 into outputs — again, of 0 or 1. Depending on their basic operation, these gates are given names such as AND, OR and NOR. Computation is also fundamental to numerous biological functions, from information processing by neural networks to nutrient sensing by microbes. In this issue, Tamsir *et al.*¹ and Regot *et al.*² describe an innovative multicellular strategy for engineering complex logic circuits, which can potentially respond to combinations of biological signals by generating a useful read-out.

In biological systems, cellular networks, which can often be thought of as assemblies of logic gates, underlie computation. To perform logic operations in such systems, researchers can engineer synthetic circuits in which biological substrates such as DNA, RNA and proteins are used as inputs, outputs and the information-processing hardware³ (Fig. 1a) in place of electronic components such as transistors and diodes. Compared with the assembly of their silicon-based counterparts, however, construction of complex biological circuits by the layering of elementary

gates is tremendously challenging⁴. This is due, in part, to crosstalk between cellular elements, the propagation of noise through networks⁵ and the metabolic burden that accompanies the expression of foreign genes in host cells^{6,7}.

The studies of Tamsir *et al.*¹ and Regot *et al.*², although different in their circuit implementation, share a common, crucial design concept: the compartmentalization of each elementary logic gate in a single cell (Fig. 1b). Here, each cell type is defined by the dedicated logic operation that it performs on inputs. And, for information flow, upstream gates produce signalling molecules that can diffuse across space into receiver cells, where these chemical 'wires' act on the downstream gate. Although a previous paper⁸ has used cellular compartmentalization in a synthetic predator–prey ecosystem, the current studies exploit it systematically.

Specifically, Tamsir and colleagues' work on the bacterium *Escherichia coli* (page 212) exploits quorum sensing — the process by which bacteria regulate their gene expression according to the local population density. The existence of multiple quorum-sensing systems with minimal crosstalk allows concurrent communication between more than one pair of sender–receiver *E. coli* populations. Regot *et al.* (page 207), meanwhile, use yeast pheromones (α -factors), taking advantage of the species-specific nature of their activity, which permits 'secure' communication in

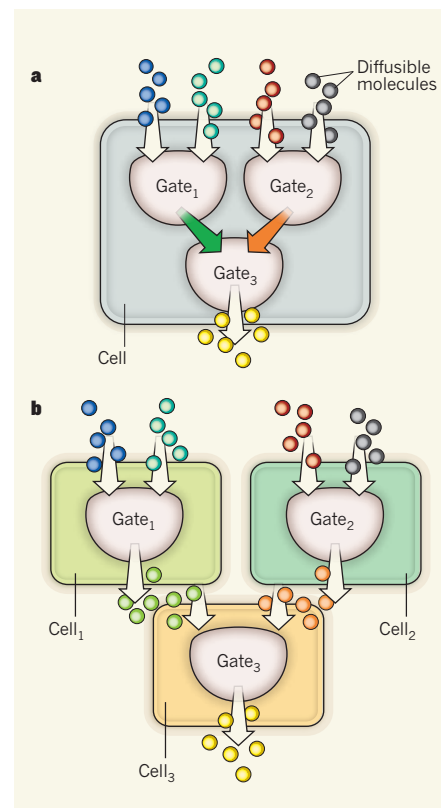


Figure 1 | A multicellular approach to implementing complex logic circuits. **a**, Previously, systems of logic gates composed of DNA, RNA and proteins have been constructed within a single cell. **b**, New work^{1,2} compartmentalizes logic gates within different cells and uses diffusible output molecules to 'wire' such cell consortia together. (Adapted from a figure by Jeffrey Wong.)

cultures containing several yeast strains.

Cellular compartmentalization is conceptually appealing. It conceals the implementation details of each encapsulated logic gate, which can be individually designed and optimized. As such, it can facilitate circuit implementation and reduce interference with the host cell's physiology by minimizing the number of components introduced into each cell strain. Therefore, to assemble a complex multicellular circuit, the experimenter needs to be concerned with only two factors: the input–output function of each cellular gate and the output–input matching between layers.

Encapsulation also physically insulates intracellular components of different gates, allowing them to be reused. To make an analogy to computer programming, each fully integrated circuit can be considered as a program. The cellular gates correspond to subroutines in the program, each dedicated to a specific function. Meanwhile, the individual components, such as promoters and transcription factors, are analogous to 'local' variables in the subroutines: the same variable can assume different roles in different subroutines without mutual dependency, thus lowering the demand for variable diversity. This design feature offers

a fundamental advantage over single-cell-based circuits⁹ (Fig. 1a), in which each regulator is analogous to a 'global' variable, accessible to all parts of the code, hence increasing the constraints on its function.

Another benefit of multicellular computing is that it allows the suppression of noise in each layer. Because the wiring-molecule output from each gate is mixed and represents the sum over a population, spurious or 'noisy' responses within a small proportion of cells can be filtered out in subsequent layers.

Certain logic gates are functionally complete: layering a sufficient number of such gates can lead to any computational operation. Illustrating this idea, both teams^{1,2} impressively exhaust all possible logic functions with no more than two inputs and one output. (Regot *et al.* also illustrate some simple three-input-one-output and two-input-two-output computations.) An intriguing question, therefore, is whether judicious choice of elementary gates to serve as logic bases might allow a reduction in the complexity of a system with respect to the number of gates or wires.

Through theoretical analysis, Regot *et al.* demonstrate that building arbitrary logic functions using the basic operations of N-IMPLIES and AND, rather than NOR, would enhance the wiring efficiency — at least in the regime of a low number of wiring molecules. The authors further alleviate the bottleneck on the biochemical diversity of these communication signals by distributing the production of final output to multiple cell populations, thereby reducing the number of computational layers.

These complementary studies indicate that it might be possible to standardize logic modules and reuse the same logic blocks to perform robust logic computations. This multicellular framework can, in principle, incorporate any regulatory components (such as DNA or RNA devices) that can be interfaced with appropriate chemical wires, thus harnessing their unique strengths. One example is the RNA aptamer system, which can be triggered by intracellular proteins to regulate gene expression¹⁰. This system could serve as a peripheral of the logic gates to respond to and incorporate versatile, user-defined inputs.

What's more, a microfluidic platform — analogous to an electronic breadboard, on which electronic components can be inserted at specific locations to achieve desired circuit connectivity — could impose physical constraints on the flow of chemical signals between cellular gates to ensure directionality and specificity, permitting the reuse of wiring molecules. Doing so could potentially reduce external intervention and increase the autonomy of circuit operation.

Nonetheless, challenges remain for the multicellular computing approaches described here. Because communication signals are produced by growing cells, for example, changes in cell densities may affect the strength of

communication. This, together with the absence of feedback mechanisms between gates, could present a major hurdle for signal correction and modulation. Moreover, in comparison with single-cell-based gates, reliance on chemically mediated communication will necessarily lead to increased time delays in the operation.

Although these issues may limit the applicability of the authors' cellular logic gates in digital computation, their design concept^{1,2} may find a home in applications in which robust circuit function is more important than speed. For instance, division of labour may be useful in metabolic engineering, in which intermediate metabolites are produced by distinct cell populations and serve simultaneously as wiring molecules. In this case, the fruits of the divided labour could be a useful chemical or protein product¹¹, whose synthesis is collectively carried out by the chemically wired populations. ■

CELL SIGNALLING

Binding the receptor at both ends

G-protein-coupled receptors initiate a wide range of signalling pathways in cells. It seems that both a G protein and an agonist molecule must bind to the receptors to persistently activate them. [SEE ARTICLE P.175](#) & [LETTERS P.236](#) & [P.241](#)

STEPHEN R. SPRANG

G-protein-coupled receptors (GPCRs) constitute a large family of proteins — consisting of more than 800 members in humans — that are embedded in the plasma membranes of most cells. They bind specifically to a diverse range of 'agonist' molecules, which convey sensory information or communicate changes in physiological states¹. Once stimulated by agonists, GPCRs bind to and activate G proteins, which are attached to the inner surface of cell membranes. The G proteins then initiate signal cascades that control a wide variety of important biological processes. In this issue, two papers from Kobilka and colleagues^{2,3} and another from the groups of Tate and Schertler⁴ report structural, biophysical and computational studies that provide insight into GPCR activation by agonists.

The membrane-embedded domains of GPCRs consist of a bundle of seven transmembrane (TM) α -helices⁵ (Fig. 1). Agonists bind to a cavity girded by four of these helices near the extracellular surface of the receptor. Agonist-induced structural changes are propagated through the transmembrane region of the receptor by rotation of certain TM helical

Bochong Li and Lingchong You are in the Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, USA. L.Y. is also at the Center for Systems Biology and Institute for Genome Sciences and Policy, Duke University. e-mail: you@duke.edu

1. Tamsir, A., Tabor, J. J. & Voigt, C. A. *Nature* **469**, 212–215 (2011).
2. Regot, S. *et al.* *Nature* **469**, 207–211 (2011).
3. Benenson, Y. *Mol. Biosyst.* **5**, 675–685 (2009).
4. Purnick, P. E. M. & Weiss, R. *Nature Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
5. Blake, W. J., Kærn, M., Cantor, C. R. & Collins, J. J. *Nature* **422**, 633–637 (2003).
6. Klumpp, S., Zhang, Z. & Hwa, T. *Cell* **139**, 1366–1375 (2009).
7. Tan, C., Marguet, P. & You, L. *Nature Chem. Biol.* **5**, 842–848 (2009).
8. Balagaddé, F. K. *et al.* *Mol. Syst. Biol.* **4**, 187 (2008).
9. Hunziker, A. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 12998–13003 (2010).
10. Culler, S. J., Hoff, K. G. & Smolke, C. D. *Science* **330**, 1251–1255 (2010).
11. Brenner, K., You, L. & Arnold, F. H. *Trends Biotechnol.* **26**, 483–489 (2008).

segments, to open a G-protein binding site on the intracellular surface. One class of GPCRs are the β -adrenergic receptors, which activate a G protein known as G_s . Activation of G_s initiates the 'fight-or-flight' response by increasing the force of cardiac contraction (causing a pounding heart), relaxing smooth muscle and mobilizing glucose in skeletal muscle.

To a first approximation, GPCRs are dynamic proteins that switch between an inactive (R) state and an active (R*) conformation that can engage G proteins. Agonists drive the conformational equilibrium towards R*; partial agonists do the same, but less efficiently, whereas inverse agonists drive the receptor towards the R state. In the absence of any such molecules, the equilibrium is biased towards R, although a small fraction of receptors remain active.

A major challenge for receptor biochemists has been to understand how agonists and inverse agonists control the conformational equilibrium. An important advance was made with the discovery that opsin, a form of the GPCR involved in vision but lacking its activating cofactor, adopts a presumed R* conformation in which the TM5 and TM6 helices are played apart, forming a cavity that accepts

the carboxy terminus of a G protein known as transducin^{6,7}. But attempts to crystallize agonist-bound R* states for other, more flexible GPCRs have met with frustration: crystal structures of the β -adrenergic receptors β_1 AR and β_2 AR have been obtained only for complexes of the receptors bound to inverse agonists^{8,9}.

Until now. In the first of the three papers published in this issue, Rasmussen *et al.*² (page 175) report that when llamas are inoculated with phospholipid vesicles that contain agonist-bound β_2 ARs, they produce antibodies that selectively bind to the R* state of the receptors. The authors identified a fragment of these antibodies — the ‘nanobody’ Nb80 — that binds to the R* state, and found that it has remarkably similar properties to those of G_s. Specifically, the affinity of β_2 AR for an agonist (isoprenaline) increases nearly a hundredfold in the presence of Nb80, indicating that the nanobody is an activator of agonist binding, like G_s. The authors did not see this effect with inverse agonists.

Using techniques honed in earlier studies⁸, Rasmussen *et al.*² obtained crystals of β_2 AR (modified in such a way as to promote crystallization) in a lipid-like environment, bound to both Nb80 and a potent β_2 AR agonist. The medium-resolution (3.5-Å) structure revealed a conformation similar to that observed for opsin^{6,7}, but with a polypeptide loop of Nb80 inserted into the cavity that, in opsin, binds the carboxy-terminal fragment of transducin. By comparing their crystal structure² with that in which β_2 AR is bound to an inverse agonist⁹, the authors observed that the transition between the R* and R states involves repacking of hydrophobic amino-acid residues in TM5 and TM6 at the ‘waist’ of the receptor (Fig. 1). The repacking is coupled with small rotations of TM5 and TM6 at hinge points within both helices, which cause the G-protein binding site to open. The rotations are facilitated by an evolutionarily conserved proline amino-acid residue in each helix: proline residues cannot participate in the hydrogen bonds that rigidify α -helices, and thus provide flexible hinge points for rotation about bonds that link adjacent peptides.

But the structural changes induced at the agonist binding site itself are small. So which agonist–receptor interactions are essential to receptor activation? Warne *et al.*⁴ (page 241) have addressed this question by solving four crystal structures in which a β_1 AR is bound to either an agonist or a partial agonist. Remarkably, they found that all of the complexes are in the inactive R state. Even so, both agonists and partial agonists pull TM7 and TM5 together,

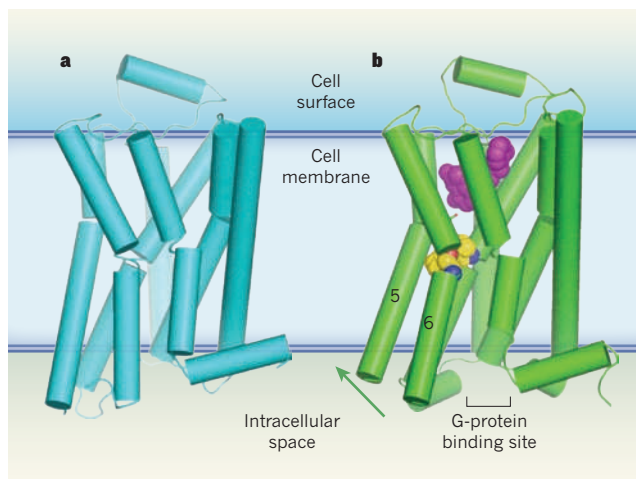


Figure 1 | Activation of G-protein-coupled receptors. G-protein-coupled receptors (GPCRs) are a family of membrane-bound proteins. They activate G proteins when bound by agonist molecules, but are inactive in the presence of inverse agonists. **a**, The structure of the β -adrenergic receptor β_2 AR — a GPCR — is shown here in its inactive (R) state⁸. The receptor is bound to the inverse agonist carazolol (not shown). Cylinders indicate α -helices. **b**, Rasmussen *et al.*² report the crystal structure of β_2 AR in complex with an agonist (purple) and with an antibody fragment (not shown), both of which stabilize the receptor's active (R*) state. The antibody binds at the G-protein binding site, and acts as a substitute for β_2 AR's G protein (G_s). The transition from the R to the R* state, induced by agonist binding, probably results from a contraction of the agonist binding site. The contraction promotes changes in the packing of amino-acid residues (yellow/blue/red spheres) between transmembrane helices 5 and 6. These changes coincide with rotation of parts of these helices (green arrow). Together, the conformational shifts create a binding site for the carboxy terminus of the G protein (or, in this case, the antibody fragment). This crystal structure², along with structures and molecular simulations in two other papers^{3,4}, suggest that both a G protein and an agonist must bind to GPCRs to stabilize the R* state.

slightly constricting the agonist binding site — an effect also seen in Rasmussen and colleagues' β_2 AR–agonist–Nb80 structure², but not in complexes of GPCRs with inverse agonists^{8–10}. Furthermore, Warne *et al.*⁴ observed that full agonists were able to form a hydrogen bond to a conserved serine amino-acid residue in TM5. The formation of this interaction seems to be the tipping point that precipitates receptor activation, and may reduce the energy barrier to the residue repacking and helix rotation that is observed in the β_2 AR–agonist–Nb80 complex².

Rosenbaum *et al.*³ (page 236) describe their attempt to trap an agonist-activated R* conformation by covalently anchoring a potent agonist to β_2 AR. When tethered to its binding site, the agonist cannot dissociate from the receptor and efficiently activates G_s — even in the presence of an inverse agonist that can displace the analogous non-covalently tethered agonist from its complex with β_2 AR. Nevertheless, the authors found that the resulting agonist– β_2 AR adduct crystallizes in the R state.

The caveat here³, as in Warne and colleagues' β_1 AR structure⁴, is that the lipid components used as surrogates for membranes in the crystals may not be as effective as native membrane bilayers in stabilizing the R* state when only agonist is present. Rosenbaum *et al.*³ therefore

performed very long molecular-dynamics simulations — extending to 30 microseconds, a record for simulations of transmembrane proteins — of the agonist-bound receptor from Rasmussen and colleagues' study², but with the nanobody removed. These simulations suggest that agonist-bound R* states of the receptor are intrinsically unstable in the absence of a G protein.

The three papers^{2–4} in this issue reveal that persistent activation of a GPCR is achieved through the binding of both an agonist and a G protein at opposite ends of the receptors relative to the lipid bilayer, where the combined binding interactions reduce the energy barriers to the formation of the R* state. Nevertheless, the conformational range and dynamics of GPCRs differ from one receptor to another. Opsin, for example, is locked completely in the R state until it is fully converted to the R* state by the light-induced isomerization of its cofactor; this accounts in part for the very high signal-to-noise ratio of visual signaling. On the other hand, β_2 ARs have constitutive activity (caused by the small fraction of the receptors that exist in the R* state at equilibrium) that may confer resistance to coronary artery disease¹¹. The instability

of R* in the absence of G proteins might also slow down agonist-dependent desensitization, a process in which prolonged exposure to agonists inactivates receptor–G-protein signaling. It will therefore be crucial to discover the molecular determinants of receptor dynamics and conformational equilibria, and how these are so finely tuned by receptor ligands. ■

Stephen R. Sprang is at the Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, Montana 59812, USA.
e-mail: stephen.sprang@umontana.edu

- Pierce, K. L., Premont, R. T. & Lefkowitz, R. J. *Nature Rev. Mol. Cell Biol.* **3**, 639–650 (2002).
- Rasmussen, S. G. F. *et al. Nature* **469**, 175–180 (2011).
- Rosenbaum, D. M. *et al. Nature* **469**, 236–240 (2011).
- Warne, T. *et al. Nature* **469**, 241–244 (2011).
- Rosenbaum, D. M., Rasmussen, S. G. F. & Koblika, B. K. *Nature* **459**, 356–363 (2009).
- Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. *Nature* **454**, 183–187 (2008).
- Scheerer, P. *et al. Nature* **455**, 497–502 (2008).
- Cherezov, V. *et al. Science* **318**, 1258–1265 (2007).
- Rosenbaum, D. M. *et al. Science* **318**, 1266–1273 (2007).
- Warne, T. *et al. Nature* **454**, 486–491 (2008).
- Piscione, F. *et al. J. Am. Coll. Cardiol.* **52**, 1381–1388 (2008).

Structure of a nanobody-stabilized active state of the β_2 adrenoceptor

Søren G. F. Rasmussen^{1,2*}, Hee-Jung Choi^{1,3*}, Juan Jose Fung^{1*}, Els Pardon^{4,5}, Paola Casarosa⁶, Pil Seok Chae⁷, Brian T. DeVree⁸, Daniel M. Rosenbaum¹, Foon Sun Thian¹, Tong Sun Kobilka¹, Andreas Schnapp⁶, Ingo Konetzki⁶, Roger K. Sunahara⁸, Samuel H. Gellman⁷, Alexander Pautsch⁶, Jan Steyaert^{4,5}, William I. Weis^{1,3} & Brian K. Kobilka¹

G protein coupled receptors (GPCRs) exhibit a spectrum of functional behaviours in response to natural and synthetic ligands. Recent crystal structures provide insights into inactive states of several GPCRs. Efforts to obtain an agonist-bound active-state GPCR structure have proven difficult due to the inherent instability of this state in the absence of a G protein. We generated a camelid antibody fragment (nanobody) to the human β_2 adrenergic receptor (β_2 AR) that exhibits G protein-like behaviour, and obtained an agonist-bound, active-state crystal structure of the receptor–nanobody complex. Comparison with the inactive β_2 AR structure reveals subtle changes in the binding pocket; however, these small changes are associated with an 11 Å outward movement of the cytoplasmic end of transmembrane segment 6, and rearrangements of transmembrane segments 5 and 7 that are remarkably similar to those observed in opsin, an active form of rhodopsin. This structure provides insights into the process of agonist binding and activation.

GPCRs activated by diffusible ligands have a spectrum of functional states¹. A GPCR may activate more than one G protein isoform or a G-protein-independent pathway such as arrestin. In the absence of a ligand, many GPCRs exhibit some basal, agonist independent activity towards one or more of these signalling pathways. Orthosteric ligands (compounds that occupy the native hormone-binding pocket) are classified according to their efficacy, that is, the effect that they have on receptor signalling through a specific pathway. Inverse agonists inhibit basal activity whereas agonists maximally activate the receptor. Partial agonists induce submaximal activity, even at saturating concentrations. Neutral antagonists have no effect on basal activity, but sterically block the activity of other ligands. Moreover, the efficacy profile of ligands for a given GPCR can differ for different downstream signalling pathways. The presence of some activity in the unliganded receptor implies low energy barriers between functional states, such that thermal fluctuations significantly sample activating conformations, and ligands with distinct efficacy profiles act by stabilizing distinct subsets of conformations.

We know little about the structural basis for the functional versatility of GPCRs. Only rhodopsin has been crystallized in different conformational states^{2–5}. The first structures of rhodopsin covalently bound to 11-*cis*-retinal represent a completely inactive state with virtually no basal activity⁵. Structures of opsin, the ligand-free form of rhodopsin, obtained from crystals grown at pH 5.6 are likely to represent active conformations^{2,3}, as the Fourier transform infrared (FTIR) spectrum of opsin at acidic pH resembles that of metarhodopsin II, the light-activated form of rhodopsin⁶. For rhodopsin, the light-induced transition from the inactive to the active state is very efficient. Rhodopsin is activated by photoisomerization of a covalent ligand, with efficient transfer of energy from the absorbed photon to the receptor. Crystal structures of low-pH opsin reveal that the protein conformation

is the same in the presence or absence of a peptide from the alpha subunit of transducin (Gt), its cognate G protein, consistent with the notion that metarhodopsin II can adopt a fully active conformation in the absence of Gt.

The crystal structures of GPCRs activated by diffusible ligands, including the human β_2 AR^{7–10}, the avian β_1 AR¹¹, and the human adenosine A_{2A} receptor¹², represent inactive conformations bound by inverse agonists. Unlike the activation of rhodopsin by light, agonists are much less efficient at stabilizing the active state of the β_2 AR, making it difficult to capture this state in a crystal structure. Fluorescence lifetime studies show that even saturating concentrations of the full agonist isoproterenol do not stabilize a single active conformation¹³. This may be due to the relatively low affinity and rapid rates of association and dissociation for β_2 AR agonists. However, in a companion manuscript we show that, even when bound to a covalent agonist, the β_2 AR crystallizes in an inactive conformation¹⁴. Experiments using a β_2 AR labelled with a conformationally sensitive fluorescent probe show that stabilization of the active state requires both agonist and Gs, the stimulatory G protein for adenylyl cyclase¹⁵. Efforts to obtain an agonist-GPCR-G protein complex are of great importance; however, this is a particularly difficult endeavour due to the biochemical challenges in working with both GPCRs and G proteins, and the inherent instability of the complex in detergent solutions. As an alternate approach, we developed a binding protein that preferentially binds to and stabilizes an active conformation, acting as a surrogate for Gs.

Nanobody-stabilized β_2 AR active state

The active G protein coupled state of the β_2 AR (and many other family A GPCRs) has characteristic functional properties. Agonists promote Gs binding to the β_2 AR and G protein binding to the receptor increases agonist affinity. We identified a camelid antibody fragment

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 279 Campus Drive, Stanford, California 94305, USA. ²Department of Neuroscience and Pharmacology, The Panum Institute, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen N, Denmark. ³Department of Structural Biology, Stanford University School of Medicine, 299 Campus Drive, Stanford, California 94305, USA. ⁴Department of Molecular and Cellular Interactions, Vlaams Instituut voor Biotechnologie (VIB), Vrije Universiteit Brussel, B-1050 Brussels, Belgium. ⁵Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium. ⁶Boehringer Ingelheim Pharma GmbH & Co. KG, Germany. ⁷Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA.

⁸Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.

*These authors contributed equally to this work.

that exhibits G protein-like behaviour towards the β_2 AR. Tylopoda (camels, dromedaries and llamas) have developed a unique class of functional antibody molecules that are devoid of light chains¹⁶. A nanobody (Nb) is the recombinant minimal-sized intact antigen-binding domain of such a camelid heavy chain antibody and is approximately 25% the size of a conventional Fab fragment. To generate receptor-specific nanobodies, a llama was immunized with purified agonist-bound β_2 AR reconstituted at high density into phospholipid vesicles. A library of single-chain nanobody clones was generated and screened against agonist bound receptor. We identified seven clones that recognized agonist-bound β_2 AR. Of these, Nb80 was chosen because it showed G-protein-like properties upon binding to both wild-type β_2 AR and β_2 AR-T4L, the β_2 AR-T4 lysozyme fusion protein used to obtain the high-resolution inactive state crystal structure^{7,9}.

We compared the effect of Nb80 with Gs on β_2 AR structure and agonist binding affinity. β_2 AR was labelled at the cytoplasmic end of transmembrane helix 6 (TM6) at Cys 265 with the fluorophore monobromobimane and reconstituted into high-density lipoprotein (HDL) particles. TM6 moves relative to TM3 and TM5 upon agonist activation (Fig. 1a), and we have shown previously that the environment around bimane covalently linked to Cys 265 changes with both agonist binding and G protein coupling, resulting in a decrease in fluorescence intensity and a red shift in λ_{\max} ¹⁵. As shown in Fig. 1b, the catecholamine agonist isoproterenol and Gs both stabilize an active-like conformation, but the effect of Gs is greater in the presence of isoproterenol, consistent with the cooperative interactions of agonist and Gs on β_2 AR structure. Nb80 alone has an effect on bimane fluorescence and λ_{\max} of unliganded β_2 AR that is similar to that of Gs (Fig. 1c). This effect was not observed in β_2 AR bound to the inverse agonist ICI-118,551. The effect of Nb80 was increased in the presence of 10 μ M isoproterenol. These results show that Nb80 does not recognize the inactive conformation of the β_2 AR, but binds efficiently to

agonist-occupied β_2 AR and produces a change in bimane fluorescence that is indistinguishable from that observed in the presence of Gs and isoproterenol.

Figure 1d and e shows the effect of Gs and Nb80 on agonist affinity for β_2 AR. β_2 AR was reconstituted into HDL particles and agonist competition binding experiments were performed in the absence or presence of Nb80 and Gs. In the absence of either protein, isoproterenol has an inhibition constant (K_i) of 107 nM. In the presence of Gs two affinity states are observed, because not all of the β_2 AR is coupled to Gs. In the Gs-coupled state the affinity of isoproterenol increases by 100-fold ($K_i = 1.07$ nM) (Fig. 1d and Supplementary Table 1). Similarly, in the presence of Nb80 the affinity of isoproterenol increases by 95-fold ($K_i = 1.13$ nM) (Fig. 1e and Supplementary Table 1). In contrast, Nb80 had little effect on β_2 AR binding to the inverse agonist ICI-118,551 (Supplementary Fig. 1 and Supplementary Table 1). These binding data indicate that Nb80 stabilizes a conformation in wild-type β_2 AR that is very similar to that stabilized by Gs, such that the energetic coupling of agonist and Gs binding is faithfully mimicked by Nb80.

The high-resolution structure of the inactive state of the β_2 AR was obtained with a β_2 AR-T4L fusion protein. We showed previously that β_2 AR-T4L has a higher affinity for isoproterenol than wild-type β_2 AR⁷. Nevertheless, in the presence of Nb80 the affinity increased by 60-fold, resulting in an affinity ($K_i = 0.56$ nM) comparable to that of wild-type β_2 AR bound to Nb80 (Fig. 1f and Supplementary Table 1). Although we cannot study G protein coupling in β_2 AR-T4L due to steric hindrance by T4L, the results show that T4L does not prevent binding of Nb80, and the nearly identical K_i values for agonist binding to wild-type β_2 AR and β_2 AR-T4L in the presence of Nb80 indicate that Nb80 stabilizes a similar conformation in these two proteins. The most likely explanation for the ability of Nb80 to bind to β_2 AR-T4L whereas Gs does not is the difference in size of these two proteins. Nb80 is approximately 14 kDa whereas the Gs heterotrimer is approximately 90 kDa.

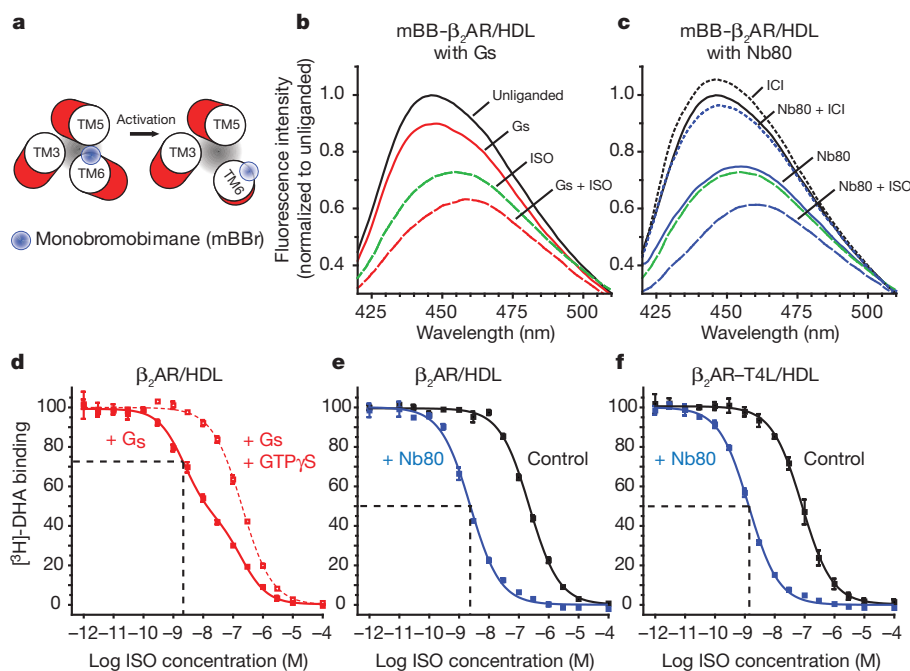


Figure 1 | Effect of Nb80 on β_2 AR structure and function. **a**, The cartoon illustrates the movement of the environmentally-sensitive bimane probe attached to Cys 265^{6,27} in the cytoplasmic end of TM6 from a more buried, hydrophobic environment to a more polar, solvent-exposed position during receptor activation that results in a decrease in fluorescence in Fig. 1b–c and Supplementary Fig. 2c, d. **b**, **c**, Fluorescence emission spectra showing ligand-induced conformational changes of monobromobimane-labelled β_2 AR reconstituted into high density lipoprotein particles (mBB- β_2 AR/HDL) in the absence (black solid line) or presence of full agonist isoproterenol (ISO, green

wide dashed line), inverse agonist ICI-118,551 (ICI, black dashed line), Gs heterotrimer (red solid line), nanobody-80 (Nb80, blue solid lines), and combinations of Gs with ISO (red wide dashed line), Nb80 with ISO (blue wide dashed line), and Nb80 with ICI (blue dashed line). **d–f**, Ligand binding curves for ISO competing against [³H]-dihydroalprenolol ([³H]-DHA) for **d**, β_2 AR/HDL reconstituted with Gs heterotrimer in the absence or presence GTP γ S; **e**, β_2 AR/HDL in the absence and presence of Nb80; and **f**, β_2 AR-T4L/HDL in the absence and presence of Nb80. Error bars represent standard errors.

High affinity β_2 AR agonist

To stabilize further the active state of the β_2 AR, we screened over 50 commercial and proprietary β_2 AR ligands. Of these, BI-167107 (Boehringer Ingelheim) had the most favourable efficacy, affinity and off-rate profile. BI-167107 is a full agonist that binds to the β_2 AR with a dissociation constant K_d of 84 pM (Supplementary Fig. 2a and b). As shown in Supplementary Fig. 2c and d, BI-167107 induces a larger change in the fluorescence intensity and λ_{\max} of bimane bound to Cys 265 than does the agonist isoproterenol. Moreover, the rate of dissociation of BI-167107 was extremely slow. Displacement of BI-167107 with an excess of the neutral antagonist alprenolol required 150 h to complete, compared with 5 s for isoproterenol.

Crystallization of β_2 AR–T4L–Nb80 complex

The β_2 AR was originally crystallized bound to the inverse agonist carazolol using two different approaches. The first crystals were obtained from β_2 AR bound to a Fab fragment that recognized an epitope composed of the amino and carboxyl terminal ends of the third intracellular loop connecting TMs 5 and 6 (ref. 8). In the second approach, the third intracellular loop was replaced by T4 lysozyme (β_2 AR–T4L)⁷. Efforts to crystallize β_2 AR–Fab complex and β_2 AR–T4L bound to BI-167107 and other agonists failed to produce crystals of sufficient quality for structure determination. We therefore attempted to crystallize BI-167107 bound to β_2 AR and β_2 AR–T4L

in complex with Nb80. Although crystals of both complexes were obtained in lipid bicelles and lipidic cubic phase (LCP), high-resolution diffraction was only obtained from crystals of β_2 AR–T4L–Nb80 grown in LCP. These crystals grew at pH 8.0 in 39–44% PEG400, 100 mM Tris, 4% DMSO and 1% 1,2,3-heptanetriol.

A merged data set at 3.5 Å was obtained from 23 crystals (Supplementary Table 2). The structure was solved by molecular replacement using the structure of the carazolol-bound β_2 AR and a nanobody as search models. Supplementary Fig. 3a shows the packing of the β_2 AR–T4L–Nb80 complex in the crystal lattice. The receptor has interactions with lattice neighbours in several directions, and is relatively well ordered (Supplementary Fig. 3a and b), with readily interpretable electron density for most of the polypeptide. Nb80 binds to the cytoplasmic end of the β_2 AR, with the third complementarity-determining region (CDR) loop projecting into the core of the receptor (Fig. 2a, and Supplementary Fig. 4).

Agonist-stabilized changes in the β_2 AR

Figure 2 b–d compares the inactive β_2 AR structure (from the carazolol-bound β_2 AR–T4L structure) with the agonist-bound β_2 AR component of the β_2 AR–T4L–Nb80 complex. The largest differences are found at the cytoplasmic face of the receptor, with outward displacement of TM5 and TM6 and an inward movement of TM7 and TM3 in the β_2 AR–T4L–Nb80 complex relative to the inactive structure. There

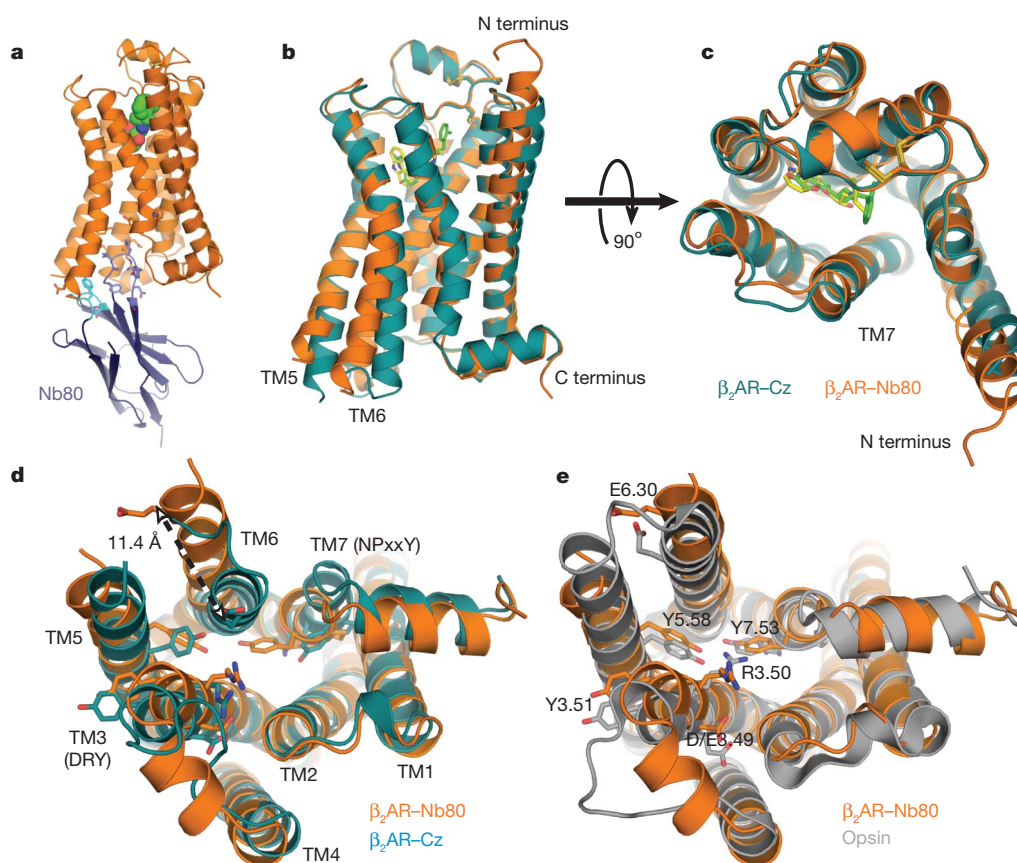


Figure 2 | Comparison of the agonist-Nb80 stabilized crystal structures of the β_2 AR with inverse agonist bound β_2 AR and opsin. The structure of inverse agonist carazolol-bound β_2 AR–T4L (β_2 AR–Cz) is shown in blue with the carazolol in yellow. The structure of BI-167107 agonist-bound and Nb80-stabilized β_2 AR–T4L (β_2 AR–Nb80) is shown in orange with BI-167107 in green. These two structures were aligned using the PyMOL align function. **a**, Side view of the β_2 AR–Nb80 complex with β_2 AR in orange and CDRs of Nb80 in light blue (CDR1) and blue (CDR3). **b**, Side view of the superimposed structures showing significant structural changes in the intracellular and G protein facing part of the receptors. **c**, Comparison of the extracellular ligand

binding domains showing modest structural changes. **d**, Cytoplasmic view showing the ionic lock interaction between Asp 3.49 and Arg 3.50 of the DRY motif in TM3 is broken in the β_2 AR–Nb80 structure. The intracellular end of TM6 is moved outward and away from the core of the receptor. The arrow indicates an 11.4 Å change in distance between the α -carbon of Glu 6.30 in the structures of β_2 AR–Cz and β_2 AR–Nb80. The intracellular ends of TM3 and TM7 move towards the core by 4 and 2.5 Å, respectively, while TM5 moves outward by 6 Å. **e**, The β_2 AR–Nb80 structure superimposed with the structure of opsin crystallized with the C-terminal peptide of Gt (transducin)². PyMOL (<http://www.pymol.org>) was used for the preparation of all structure figures.

are relatively small changes in the extracellular surface (Fig. 2c). The second intracellular loop (ICL2) between TM3 and TM4 adopts a two-turn alpha helix (Fig. 2d), similar to that observed in the turkey β_1 AR structure¹¹. The absence of this helix in the inactive β_2 AR structure may reflect crystal lattice contacts involving ICL2.

Figure 2a and Supplementary Fig. 4a–c show details of interaction of Nb80 with the cytoplasmic side of the β_2 AR. An eight-amino-acid sequence of CDR3 penetrates into a hydrophobic pocket formed by amino acids from TM segments 3, 5, 6 and 7. A four-amino-acid sequence of CDR1 provides additional stabilizing interactions with cytoplasmic ends of TM segments 5 and 6. CDR3 occupies a position similar to the carboxyl terminal peptide of transducin in opsin² (Supplementary Fig. 4c, d). The majority of interactions between Nb80 and the β_2 AR are mediated by hydrophobic contacts.

When comparing the agonist- and inverse agonist-bound structures, the largest change is observed in TM6, with an 11.4-Å movement of the helix at Glu 268^{6.30} (part of the ionic lock) (superscripts in this form indicate Ballesteros–Weinstein numbering for conserved GPCR residues¹⁷) (Fig. 2d). This large change is effected by a small clockwise rotation of TM6 in the turn preceding the conserved Pro 288^{6.50}, enabled by the interrupted backbone hydrogen bonding at the proline and repacking of Phe 282^{6.44} (see below), which swings the helix outward.

The changes in agonist-bound β_2 AR–T4L–Nb80 relative to the inactive carazolol-bound β_2 AR–T4L are remarkably similar to those

observed between rhodopsin and opsin^{2,3} (Fig. 2e). The salt bridge in the ionic lock between highly conserved Arg 131^{3.50} and Asp/Glu 130^{3.49} is broken. In opsin, Arg 135^{3.50} interacts with Tyr 223^{5.58} in TM5 and a backbone carbonyl of the transducin peptide. Arg 131^{3.50} of β_2 AR likewise interacts with a backbone carbonyl of CDR3 of Nb80. However, Nb80 precludes an interaction between Arg 131^{3.50} and Tyr 219^{5.58}, even though the tyrosine occupies a similar position in opsin and agonist-bound β_2 AR–T4L–Nb80. As in opsin, Tyr 326^{7.53} of the highly conserved NPxxY sequence moves into the space occupied by TM6 in the inactive state. In carazolol-bound β_2 AR–T4L we observed a network of hydrogen bonding interactions involving highly conserved amino acids in TMs 1, 2, 6 and 7 and several water molecules⁷. Although the resolution of the β_2 AR–T4L–Nb80 structure is inadequate to detect water molecules, it is clear that the structural changes we observe would substantially alter this network.

In contrast to the relatively large changes observed in the cytoplasmic domains of β_2 AR–T4L–Nb80, the changes in the agonist-binding pocket are fairly subtle. Figure 3 shows a comparison of the binding pockets of the inverse agonist- and agonist-bound structures. An omit map of the ligand-binding pocket is provided in Supplementary Fig. 5. Many of the interactions between the agonist BI-167107 and the β_2 AR are similar to those observed with the inverse agonist carazolol. The alkylamine and the β -OH of both ligands form polar interactions with Asp 113^{3.32} in TM3, and with Asn 312^{7.39} and Tyr 316^{7.43} in TM7. The

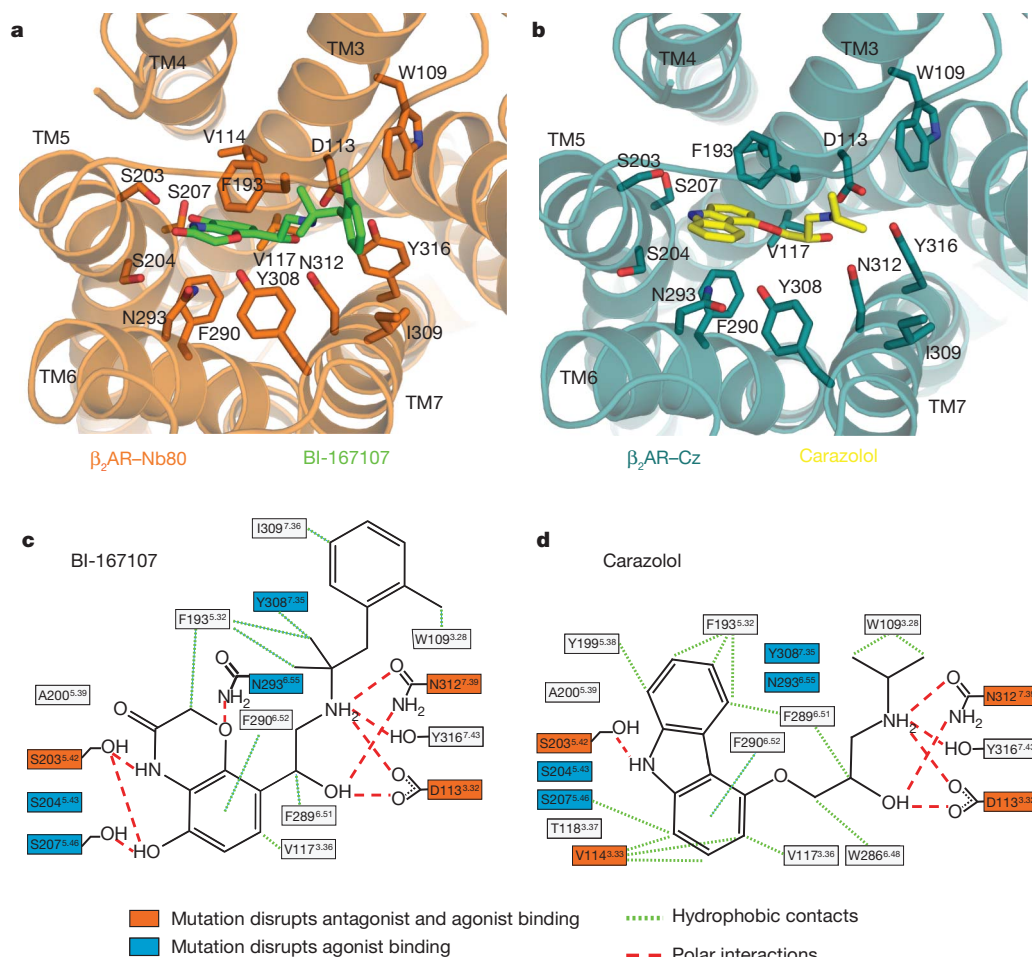


Figure 3 | Ligand binding pocket of BI-167107 and carazolol-bound β_2 AR structures. **a, b**, Extracellular views of the agonist BI-167107-bound (**a**) and carazolol-bound (**b**) structures, respectively. Residues within 4 Å of one or both ligands are shown as sticks. In all panels, red and blue represent oxygen and nitrogen, respectively. **c, d**, Schematic representation of the interactions between the β_2 AR and the ligands BI-167107 (**c**) and carazolol (**d**). The residues

shown here have at least one atom within 4 Å of the ligand in the crystal structures. Mutations of amino acids in orange boxes have been shown to disrupt both antagonist and agonist binding. Mutations of amino acids in blue boxes have been shown to disrupt agonist binding. Green lines indicate potential hydrophobic interactions and orange lines indicate potential polar interactions.

agonist has a longer alkyl substituent on the amine, which ends with a phenyl ring that lies in a hydrophobic pocket formed by Trp 109^{3,28}, Phe 193^{5,32} and Ile 309^{7,36}.

The greatest difference between inactive and active structures in the ligand-binding site is an inward bulge of TM5 centred around Ser 207^{5,46}, whose C α position shifts by 2.1 Å (Fig. 4a). In addition, there are smaller inward movements of TM6 and TM7. The basal activity shown by the β_2 AR indicates that the protein structure surrounding the binding pocket is relatively dynamic in the absence of ligand, such that it samples active and inactive conformations. The presence of Pro 211^{5,50} in the following turn, which cannot form a hydrogen bond with the backbone at Ser 207^{5,46}, is likely to lower the barrier to the transition between the conformations observed in the presence of carazolol and BI-167107. There are extensive interactions between the carbonyl oxygen, amine and hydroxyl groups on the heterocycle of BI-167107 and Ser 203^{5,42} and 207^{5,46} in TM5, as well as Asn 293^{6,55} in TM6 and Tyr 308^{7,35} in TM7. In contrast, there is only one polar interaction between the nitrogen in the heterocycle of carazolol and Ser 203^{5,42}. Interactions of Ser 203^{5,42}, Ser 204^{5,43} and Ser 207^{5,46} with catecholamine hydroxyls have been proposed, on the basis of mutagenesis studies showing that these serines are important for agonist binding and activation^{18,19}. Whereas Ser 204^{5,43} does not interact directly with the ligand, it forms a hydrogen bond with Asn 293^{6,55} on TM6, which is in turn linked to Tyr 308^{7,35} of extracellular loop 3 (ECL3) (Fig. 3a). This tyrosine packs against Phe 193^{5,32} of ECL2, and both residues move to close off the ligand-binding site from the extracellular space.

Asn 293^{6,55} contributes to enantiomeric selectivity for catecholamine agonists²⁰. The β -OH of BI-167107 does not interact with Asn 293^{6,55}, but forms hydrogen bonds with Asp 113^{3,32} and Asn 312^{7,39}, similar to what is observed for carazolol in the inactive structure. The chirality of the β -OH influences the spatial position of the aromatic ring system in β_2 AR ligands, so the effect of Asn 293^{6,55} on β -OH enantiomeric selectivity may arise from its direct interaction with the aromatic ring system of the ligand, as well as its positioning of Ser 204^{5,43} and Tyr 308^{7,35}, which also interact with this portion of the ligand. However, BI-167107 is not a catecholamine, and it is possible that the β -OH of catecholamine agonists, such as adrenaline and noradrenaline, has a direct interaction with Asn 293^{6,55}, because mutation of Asn 293^{6,55} has a stronger influence on the preference for the chirality of the β -OH of catecholamine agonists, compared with non-catechol agonists and antagonists²⁰.

Trp^{6,48} is highly conserved in Family A GPCRs, and it has been proposed that its rotameric state has a role in GPCR activation (rotamer

toggle switch)²¹. We observe no change in the side chain rotamer of Trp 286^{6,48} in TM6 (Fig. 4a), which lies near the base of the ligand-binding pocket, although its position shifts slightly in concert with rearrangements of nearby residues Ile 121^{3,40} and Phe 282^{6,44}. Although there is spectroscopic evidence for changes in the environment of Trp^{6,48} upon activation of rhodopsin²², a rotamer change is not observed in the crystal structures of rhodopsin and low-pH opsin. Moreover, recent mutagenesis experiments on the serotonin 5HT4 receptor demonstrate that Trp^{6,48} is not required for activation of this receptor by serotonin²³. These observations indicate that, although changes in hydrophobic packing alter the conformation of the receptor in this region, changes in the Trp^{6,48} rotamer do not occur as part of the activation mechanism.

It is interesting to speculate how the small changes around the agonist-binding pocket are coupled to much larger structural changes in the cytoplasmic regions of TMs 5, 6 and 7 that facilitate binding of Nb80 and Gs. A potential conformational link is shown in Fig. 4. Agonist interactions with Ser 203^{5,42} and 207^{5,46} stabilize a receptor conformation that includes a 2.1-Å inward movement of TM5 at position 207^{5,46} and 1.4-Å inward movement of the conserved Pro 211^{5,50} relative to the inactive, carazolol-bound structure. In the inactive state, the relative positions of TM5, TM3, TM6 and TM7 are stabilized by interactions between Pro 211^{5,50}, Ile 121^{3,40}, Phe 282^{6,44} and Asn 318^{7,45}. The position of Pro 211^{5,50} observed in the agonist structure is incompatible with this network of interactions, and Ile 121^{3,40} and Phe 282^{6,44} are repositioned, with a rotation of TM6 around Phe 282^{6,44} leading to an outward movement of the cytoplasmic end of TM6.

Although some of the structural changes observed in the cytoplasmic ends of transmembrane domains of the β_2 AR-T4L-Nb80 complex arise from specific interactions with Nb80, the fact that Nb80 and Gs induce or stabilize similar structural changes in the β_2 AR, as determined by fluorescence spectroscopy and by agonist binding affinity, suggests that Nb80 and Gs recognize similar agonist-stabilized conformations. The observation that the transmembrane domains of rhodopsin and the β_2 AR undergo similar structural changes upon activation provides further support that the agonist-bound β_2 AR-T4L-Nb80 represents an active conformation and is consistent with a conserved mechanism of G protein activation.

However, the mechanism by which agonists induce or stabilize these conformational changes likely differs for different ligands and for different GPCRs. The conformational equilibria of rhodopsin and β_2 AR differ, as shown by the fact that rhodopsin appears to adopt a fully active conformation in the absence of a G protein²⁴ whereas β_2 AR cannot¹⁵. Thus, the energetics of activation and conformational

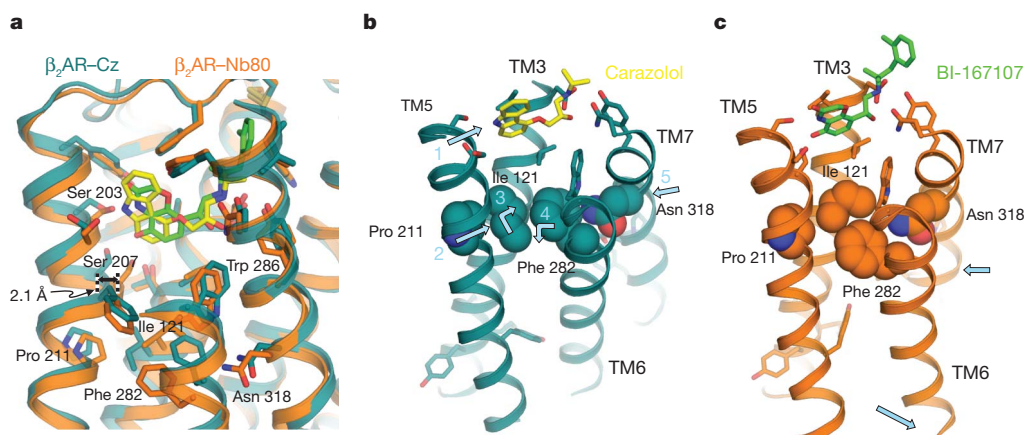


Figure 4 | Rearrangement of transmembrane segment packing interactions upon agonist binding **a**, The BI-167107- and carazolol-bound structures are superimposed to show structural differences propagating from the ligand-binding pocket. BI-167107 and carazolol are shown in green and yellow, respectively. **b**, Packing interactions that stabilize the inactive state are observed

between Pro 211 in TM5, Ile 121 in TM3, Phe 282 in TM6 and Asn 318 in TM7. **c**, The inward movement of TM5 upon agonist binding destabilizes the packing of Ile 121 and Pro 211, resulting in a rearrangement of interactions between Ile 121 and Phe 282. These changes contribute to a rotation and outward movement of TM6 and an inward movement of TM7.

sampling can differ among different GPCRs, which likely gives rise to the variety of ligand efficacies displayed by these receptors. An agonist need only disrupt one key intramolecular interaction needed to stabilize the inactive state, as constitutive receptor activity can result from single mutations of amino acids from different regions of GPCRs²⁵. Thus, disruption of these stabilizing interactions either by agonists or mutations lowers the energy barrier separating inactive and active states and increases the probability that a receptor can interact with a G protein.

METHODS SUMMARY

Crystallization. Preparation of β_2 AR–T4L and Nb80 are described in Methods. BI-167107-bound β_2 AR–T4L and Nb80 preincubated in 1:1.2 molar ratio were mixed in monoolein containing 10% cholesterol in 1:1.5 protein to lipid ratio (w/w). Initial crystallization leads were identified and optimized in 24-well glass sandwich plates using 50 nl protein:lipid drops overlaid with 0.8 μ l precipitant solution in each well and sealed with a glass cover slip. Crystals for data collection were grown at 20 °C in hanging-drop format using 0.8 μ l reservoir solution (36 to 44% PEG 400, 100 mM Tris pH 8.0, 4% DMSO, 1% 1,2,3-heptanetriol) diluted two- to fourfold in water. Crystals grew to full size, typically $40 \times 5 \times 5 \mu\text{m}^3$, within 7 to 10 days. Crystals were flash-frozen and stored in liquid nitrogen with reservoir solution as cryoprotectant. Diffraction data collection and processing, and structure solution and refinement are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 July; accepted 1 November 2010.

- Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
- Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
- Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).
- Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).
- Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* [see comments] **289**, 739–745 (2000).
- Vogel, R. & Siebert, F. Conformations of the active and inactive states of opsin. *J. Biol. Chem.* **276**, 38487–38493 (2001).
- Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into β -adrenergic receptor function. *Science* **318**, 1266–1273 (2007).
- Rasmussen, S. G. *et al.* Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
- Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2 -adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
- Hanson, M. A. *et al.* A specific cholesterol binding site is established by the 2.8 Å structure of the human β_2 -adrenergic receptor. *Structure* **16**, 897–905 (2008).
- Warne, T. *et al.* Structure of a β_1 -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
- Jaakola, V. P. *et al.* The 2.6 Å crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
- Ghanouni, P. *et al.* Functionally different agonists induce distinct conformations in the G protein coupling domain of the β_2 adrenergic receptor. *J. Biol. Chem.* **276**, 24433–24436 (2001).
- Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist– β_2 adrenoceptor complex. *Nature* doi:10.1038/nature09665 (this issue).
- Yao, X. J. *et al.* The effect of ligand efficacy on the formation and stability of a GPCR–G protein complex. *Proc. Natl Acad. Sci. USA* **106**, 9501–9506 (2009).
- Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein coupled receptors. *Meth. Neurosci.* **25**, 366–428 (1995).
- Strader, C. D. *et al.* Identification of residues required for ligand binding to the β -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **84**, 4384–4388 (1987).
- Liapakis, G. *et al.* The forgotten serine. A critical role for Ser-203^{5,42} in ligand binding to and activation of the β_2 -adrenergic receptor. *J. Biol. Chem.* **275**, 37779–37788 (2000).
- Wieland, K., Zuurmond, H. M., Krasel, C., Ijzerman, A. P. & Lohse, M. J. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the β_2 -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **93**, 9276–9281 (1996).
- Shi, L. *et al.* β_2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J. Biol. Chem.* **277**, 40989–40996 (2002).
- Ahuja, S. & Smith, S. O. Multiple switches in G protein-coupled receptor activation. *Trends Pharmacol. Sci.* **30**, 494–502 (2009).
- Pellissier, L. P. *et al.* Conformational toggle switches implicated in basal constitutive and agonist-induced activated states of 5-hydroxytryptamine-4 receptors. *Mol. Pharmacol.* **75**, 982–990 (2009).
- Altenbach, C., Kusnetzow, A. K., Ernst, O. P., Hofmann, K. P. & Hubbell, W. L. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc. Natl Acad. Sci. USA* **105**, 7439–7444 (2008).
- Parnot, C., Miserey-Lenkei, S., Bardin, S., Corvol, P. & Clauser, E. Lessons from constitutively active mutants of G protein-coupled receptors. *Trends Endocrinol. Metab.* **13**, 336–343 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from National Institutes of Health Grants NS028471 and GM083118 (B.K.K.), GM56169 (W.I.W.), P01 GM75913 (S.H.G.), and P60DK-20572 (R.K.S.), the Mathers Foundation (B.K.K. and W.I.W.), the Lundbeck Foundation (Junior Group Leader Fellowship, S.G.F.R.), the University of Michigan Biomedical Sciences Scholars Program (R.K.S.), the Fund for Scientific Research of Flanders (FWO-Vlaanderen) and the Institute for the encouragement of Scientific Research and Innovation of Brussels (ISRI) (E.P. and J.S.).

Author Contributions S.G.F.R. screened and characterized high affinity agonists, identified and determined dissociation rate of BI-167107, screened, identified and characterized MNG-3, performed selection and characterization of nanobodies, purified and crystallized the receptor with Nb80 in LCP, optimized crystallization conditions, grew crystals for data collection, reconstituted receptor in HDL particles and determined the effect of Nb80 and Gs on receptor conformation and ligand binding affinities, assisted with data collection and preparing the manuscript. H.-J.C. processed diffraction data, solved and refined the structure, and assisted with preparing the manuscript. J.J.F. expressed, purified, selected and characterized nanobodies, purified and crystallized receptor with nanobodies in bicelles, assisted with growing crystals in LCP, and assisted with data collection. E.P. performed immunization, cloned and expressed nanobodies, and performed the initial selections. J.S. supervised nanobody production. P.S.K. and S.H.G. provided MNG-3 detergent for stabilization of purified β_2 AR. B.T.D. and R.K.S. provided ApoA1 and Gs protein, and reconstituted β_2 AR in HDL particles with Gs. D.M.R. characterized the usefulness of MNG-3 for crystallization in LCP and assisted with manuscript preparation. F.S.T. expressed β_2 AR in insect cells and with T.S.K. performed the initial stage of β_2 AR purification. A.P., A.S. assisted in selection of the high-affinity agonist BI-167107. I.K. synthesized BI-167,107. P.C. characterized the functional properties of BI-167,107 in CHO cells. W.I.W. oversaw data processing, structure determination and refinement, and assisted with writing the manuscript. B.K.K. was responsible for the overall project strategy and management, prepared β_2 AR in lipid vesicles for immunization, harvested and collected data on crystals, and wrote the manuscript.

Author Information Coordinates and structure factors for β_2 AR–Nb80 are deposited in the Protein Data Bank (accession code 3POG). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.K.K. (kobilka@stanford.edu) or W.I.W. (bill.weis@stanford.edu).

METHODS

Preparation of β 2AR–T4L and nanobody-80 for crystallography. β 2AR–T4L was expressed in Sf-9 insect cell cultures infected with β 2AR–T4L baculovirus, and solubilized according to methods described previously²⁶. Functional protein was obtained by M1 Flag affinity chromatography (Sigma) before and following alprenolol-Sepharose chromatography²⁶. In the second M1 chromatography step, receptor-bound alprenolol was exchanged for high-affinity agonist BI-167107 and dodecylmaltoside was exchanged for the MNG-3 amphiphile (11,11-bis- β -D-maltopyranosidylmethyl-heneicosane, Supplementary Fig. 6, obtained from P. S. Chae and S. H. Gellman) for increased receptor stability. The agonist-bound and detergent-exchanged β 2AR–T4L was eluted in 10 mM HEPES pH 7.5, 100 mM NaCl, 0.02% MNG-3 and 10 μ M BI-167107 followed by removal of N-linked glycosylation by treatment with PNGaseF (NEB). The protein was concentrated to \sim 50 mg ml^{−1} with a 100 kDa molecular weight cut off Vivaspinn concentrator (Vivascience).

Nanobody-80 (Nb80) bearing a carboxy-terminal His₆ tag was expressed in the periplasm of *Escherichia coli* strain WK6 following induction with IPTG. Cultures of 0.6 l were grown to $A_{600} = 0.7$ at 37 °C in TB media containing 0.1% glucose, 2 mM MgCl₂, and 50 μ g ml^{−1} ampicillin. Induced cultures were grown overnight at 28 °C. Cells were harvested by centrifugation and lysed in ice-cold buffer (50 mM Tris pH 8.0, 12.5 mM EDTA and 0.125 M sucrose), then centrifuged to remove cell debris. Nb80 was purified by nickel affinity chromatography, dialysed against buffer (10 mM HEPES pH 7.5, 100 mM NaCl), and spin concentrated to \sim 120 mg ml^{−1}.

Crystallization. BI-167107 bound β 2AR–T4L and Nb80 were mixed in 1:1.2 molar ratio, incubated 2 h at room temperature before mixing with liquefied monoolein (M7765, Sigma) containing 10% cholesterol (C8667, Sigma) in 1:1.5 protein to lipid ratio (w/w) using the twin-syringe mixing method reported previously²⁷. Initial crystallization leads were identified using in-house screens and optimized in 24-well glass sandwich plates using 50 nl protein:lipid drops manually delivered and overlaid with 0.8 μ l precipitant solution in each well and sealed with a glass cover slip. Crystals for data collection were grown at 20 °C by hanging drop vapour diffusion using 0.8 μ l reservoir solution (36 to 44% PEG 400, 100 mM Tris pH 8.0, 4% DMSO, 1% 1,2,3-heptanetriol) diluted two- to fourfold in Milli-Q water. Crystals grew to full size within 7 to 10 days. Crystals were flash-frozen and stored in liquid nitrogen with reservoir solution as cryoprotectant.

Microcrystallography data collection and processing. Diffraction data were measured at beamline 23-ID of the Advanced Photon Source, using a 10- μ m diameter beam. Low dose 1.0° rotation images were used to locate and centre crystals for data collection. Data were measured in 1.0° frames with exposure times typically 5–10 s with a 5 \times attenuated beam. Only 5–10° of data could be measured before significant radiation damage occurred. Data were integrated and scaled with the HKL2000 package²⁸.

Structure solution and refinement. Molecular replacement phases were obtained with the program Phaser²⁹. The search models were (1) the high-resolution carazolol-bound β 2AR structure, PDB ID 2RH1, but with T4L and all water, ligand and lipid molecules removed) and a nanobody (PDB ID 3DWT, water molecules removed) as search models. The rotation and translation function Z scores were 8.7 and 9.0 after placing the β 2AR model, and the nanobody model placed subsequently had rotation and translation function Z scores of 3.5 and 11.5. The model was refined in Phenix³⁰ and Buster³¹, using a group B factor model with one B for main chain and one B for side chain atoms. Refinement statistics are given in Supplementary Table 2. Despite the strong anisotropy (Supplementary Table 2), the electron density was clear for the placement of side chains.

Ligand binding on receptor reconstituted in HDL particles. The effect of Nb80 and Gs on the receptors affinity for agonists was compared in competition binding experiments. The β 2AR and β 2AR–T4L (both truncated at position 365) purified as previously described^{7,8} were reconstituted in high-density lipoprotein (HDL) particles followed by reconstitution of Gs into HDL particles containing β 2AR according to previously published methods³². [³H]-dihydroalprenolol ([³H]-DHA; 0.6 nM) was used as radioligand and agonist (−)-isoproterenol (ISO) or inverse agonist ICI-118,551 (ICI) as competitor. Nb80 was used at 1 μ M. GTP γ S was used at 10 μ M. TBS (50 mM Tris pH 7.4, 150 mM NaCl) containing 0.1% BSA was used as binding buffer. Bound [³H]-DHA was separated from unbound on a Brandel harvester by passing over a Whatman GF/B filter (presoaked in TBS with 0.3% polyethylenimine) and washed in cold TBS. Radioligand binding was measured in a Beckman LS6000 scintillation counter.

Ligand binding affinity (K_d) of DHA was determined from saturation binding curves using GraphPad Prism software. Normalized ISO competition binding data were fit to a two-site competition binding model by using GraphPad Prism. Binding affinities of ISO (K_i values, tabulated in Supplementary Table 1) were determined from 50% inhibitory concentration (IC₅₀) values using the equation $K_i = IC_{50}/(1 + [L]/K_d)$.

cAMP assay. To determine the functional potency of BI-167107, changes in intracellular cAMP levels were determined with CHO-h β 2AR cells in suspension (15,000 cells per well) by using Alphascreen technology (PerkinElmer Life and Analytical Sciences) and a 384-well plate format (Optiplate; PerkinElmer Life and Analytical Sciences), according to the manufacturer's protocol. In brief, cells were stimulated with the respective agonists at different concentrations in Hanks' buffered saline solution supplemented with 5 mM HEPES, 0.1% bovine serum albumin and 500 mM 3-isobutyl-1-methylxanthine for 30 min at room temperature. Cells were lysed by using Alphascreen reagents. After 2 h, plates were read on an Envision plate reader (PerkinElmer Life and Analytical Sciences). The concentration of cAMP in the samples was calculated from a standard curve.

Bimane fluorescence spectroscopy on β 2AR reconstituted in HDL particles. To compare the effects on receptor conformation of Gs and Nb80 binding the purified β 2AR was labelled with the environmentally sensitive fluorescent probe monobromobimane (Invitrogen) at cysteine 265 located in the cytoplasmic end of TM6, and reconstituted into HDL particles (mBB- β 2AR/HDL). Prior to obtaining fluorescence emission spectra, 10 nM mBB- β 2AR/HDL was incubated 30 min at room temperature in buffer (20 mM HEPES pH 7.5, 100 mM NaCl) in the absence or presence of 10 μ M ISO, 1 μ M ICI, 300 nM Gs heterotrimer, or 300 nM Nb80, or in combinations of ISO with Gs, ISO with Nb80, and ICI with Nb80. Fluorescence spectroscopy was performed on a Spex FluoroMax-3 spectrofluorometer (Jobin Yvon) with photon-counting mode, using an excitation and emission bandpass of 5 nm. Excitation was set at 370 nm and emission was collected from 415 to 535 nm in 1-nm increments with 0.3 nm^{−1} integration time. Fluorescence intensity was corrected for background fluorescence from buffer and ligands. The curves shown in Fig. 1b and c are each the average of triplicate experiments.

High affinity β 2AR agonist screening by bimane fluorescence spectroscopy. To obtain high affinity agonist candidates with slow dissociation rates for crystallography, a screening process of commercially available drugs and compound libraries from medicinal and biotech industry was initiated. Screening was conducted in several rounds on more than 50 compounds. Each compound (10 μ M) was incubated with 100 nM purified mBB- β 2AR in DDM buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 0.1% dodecylmaltoside (DDM)) for 30 min at room temperature before emission scanning, using same equipment and settings as described in the section above. Compounds inducing the largest red shift in λ_{max} and decrease in bimane fluorescence emission were identified. Closely related structural analogues were subsequently screened using same criteria for selection. Several lead candidate compounds were then subjected to dissociation experiments to identify the agonist with the slowest rate of dissociation. In these experiments, 100 nM mBB- β 2AR was incubated with 1 μ M lead compound in DDM buffer for 2 h at room temperature before obtaining the emission scan at $t = 0$ (example in Supplementary Fig. 2d, green spectra). An excess amount (200 μ M) of the neutral antagonist alprenolol (ALP) was added to identical samples followed by measurement of bimane emission at various time points in a period up to 7 days or until complete dissociation of agonist.

26. Kobilka, B. K. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. *Anal. Biochem.* **231**, 269–271 (1995).
27. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
30. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D* **61**, 850–855 (2005).
31. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
32. Whorton, M. R. et al. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).

A distortion of very-high-redshift galaxy number counts by gravitational lensing

J. Stuart B. Wyithe¹, Haojing Yan², Rogier A. Windhorst³ & Shude Mao^{4,5}

The observed number counts of high-redshift galaxy candidates^{1–8} have been used to build up a statistical description of star-forming activity at redshift $z \gtrsim 7$, when galaxies reionized the Universe^{1,2,9,10}. Standard models¹¹ predict that a high incidence of gravitational lensing will probably distort measurements of flux and number of these earliest galaxies. The raw probability of this happening has been estimated to be ~ 0.5 per cent (refs 11, 12), but can be larger owing to observational biases. Here we report that gravitational lensing is likely to dominate the observed properties of galaxies with redshifts of $z \gtrsim 12$, when the instrumental limiting magnitude is expected to be brighter than the characteristic magnitude of the galaxy sample. The number counts could be modified by an order of magnitude, with most galaxies being part of multiply imaged systems, located less than 1 arcsec from brighter foreground galaxies at $z \approx 2$. This lens-induced association of high-redshift and foreground galaxies has perhaps already been observed among a sample of galaxy candidates identified at $z \approx 10.6$. Future surveys will need to be designed to account for a significant gravitational lensing bias in high-redshift galaxy samples.

Along random lines of sight, the raw probability (or optical depth) of multiple imaging of objects at high redshifts—owing to gravitational lensing by individual foreground field galaxies^{11,12}—is $\sim 0.5\%$. However, all galaxy populations are observed to have a characteristic luminosity (L_*), brighter than which galaxy numbers drop exponentially, and below which numbers rise with a very steep power-law slope^{1,4,6}. The potential for gravitational lensing to modify the observed statistics therefore increases dramatically, owing to the magnification of numerous, intrinsically faint galaxies to observed fluxes that are above the survey limit. This effect, which is known as magnification bias¹³, leads to an excess of gravitationally lensed galaxies among flux-limited samples. Magnification bias is expected to be particularly significant at high redshifts ($z \gtrsim 8$), where current observations may only be probing the exponential tail of the luminosity function⁴, so that the number density could be rising very rapidly towards the detection limit. Indeed, multiply imaged candidates at $z \gtrsim 7$ have already been discovered behind foreground clusters via targeted searches^{14–16}, demonstrating this to be an efficient method for finding faint high-redshift galaxies^{14,17}.

We assess magnification bias among high-redshift galaxies assuming singular, spherical, isothermal gravitational lenses, which produce one or two images, and designate the apparent magnitude of the more magnified image (the only image in the absence of lensing) as $m_{AB,1}$. We then calculate, as a function of the assumed characteristic luminosity (expressed in terms of absolute magnitude, M_*), the fraction of galaxies brighter than the magnitude limit (m_{lim}) for the Hubble Ultra Deep Field (HUDF) that would be multiply imaged (designated F_{lens}). Such curves are shown at $z = 6, 7, 8.6$ and 10.6 in Fig. 1a. The superimposed filled and open points correspond to lens fractions for different estimates¹⁴ of M_* at these redshifts. At $z \approx 6–7$, we expect only $\sim 1\%$ of galaxies to be lensed. At $z \approx 8–10$, however, we expect a lensed fraction

of a few to a few tens of per cent, depending on the true value of M_* . Note that since current survey limits are significantly fainter than M_* at $z \approx 6–7$, the lens fraction is quite insensitive to M_* . However, at higher redshifts where the survey limits might be much closer to M_* , the lensing fraction is very sensitive to its uncertain value.

Predictions for a significant lens fraction at $z \gtrsim 8$ stand in apparent contrast to the fact that no image pairs have been identified in the HUDF. However, we find the probability that a multiply imaged galaxy, with observed $m_{AB,1}$ has a corresponding second image with $m_{AB,2} < m_{lim}$ (that is, detectable with the HUDF data) to be only $\sim 10\%$, even for galaxies that are one magnitude brighter than m_{lim} (Supplementary Information). Thus, as shown in Fig. 1b, the fraction of galaxies (F_{mult}) that are detected as multiply imaged systems in the HUDF is an order of magnitude lower than the true lensed fraction. Although this fraction would increase somewhat if elliptical lenses were included in our analysis, multiply imaged systems are not expected to be observed in the current data. On the other hand, magnification bias also leads to a concentration of high-redshift sources—both singly and multiply imaged—around foreground galaxies^{18–20}. The resulting correlation between high-redshift candidates and bright foreground galaxies therefore offers an alternative avenue to observing the effect of gravitational lensing. A schematic diagram illustrating this point, as well as magnification bias, is included as Supplementary Fig. 1.

To quantify this correlation, we first determine the distribution of separations between random lines of sight and the nearest bright ($H \leq 25$ mag) foreground galaxy in the HUDF, measured as the angular distance to the centroid. This is shown by the dotted black line in Fig. 1c. This distribution can be compared to the predictions of our model (black dashed line in Fig. 1c). If the candidate sample consists of both multiply imaged and unmagnified galaxies, then the observed distribution of separations should be a weighted sum of the random and the lensed line-of-sight distributions. The correct weighting is the probability for gravitational lensing, F_{lens} . Two examples (blue dotted and dashed lines) are shown in Fig. 1c. The fraction of galaxies found within $\Delta\theta \approx 1–2$ arcsec of a foreground galaxy is very sensitive to the characteristic luminosity if $M_* \gtrsim -19$ mag, providing a potential observable for the influence of lensing on the number counts of $z \gtrsim 8$ candidates.

For comparison with the lensing predictions, we have measured the distribution of separations between a sample of $z \approx 10.6$ candidates⁴ and their nearest bright ($H \leq 25$ mag) foreground galaxy. Comparing the distributions, we find that these candidates are observed to be closer to bright foreground galaxies than are random lines of sight. On the other hand, the candidates are found at larger separations from foreground galaxies than would be predicted if they were all multiply imaged. Quantitatively, the Kolmogorov–Smirnov probabilities between the observed distributions and the all-random model or the all-lensed model (Supplementary Information) indicate that both models are rejected at high significance. This suggests that a fraction of candidates may be gravitationally lensed. Moreover, we have generated the

¹School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia. ²Center for Cosmology and AstroParticle Physics, The Ohio State University, Columbus, Ohio 43210, USA. ³School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287-1404, USA. ⁴Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester M13 9PL, UK. ⁵National Astronomical Observatories of China, Chinese Academy of Sciences, Beijing 100012, China.

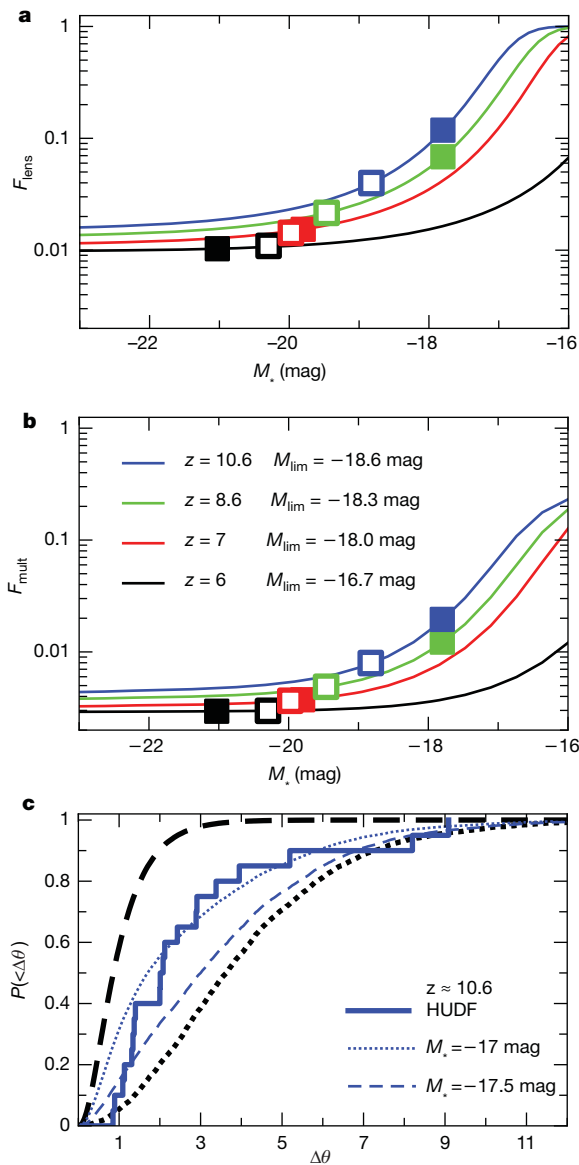


Figure 1 | Gravitational lens fractions among candidate high-redshift HUDF galaxies. **a**, The fraction of multiply imaged high-redshift galaxies (with $m_{AB,1} < m_{lim}$). **b**, The fraction of high-redshift galaxies in which multiple images could be detected in the HUDF (with $m_{AB,2} < m_{lim}$). **c**, The probability distribution of image separations (at $z \approx 10.6$) relative to the nearest bright foreground galaxy, in the cases of random lines-of-sight (black dotted line), of gravitational lenses (black dashed line), and for composite distributions computed for two (faint) values of M_* (blue dotted and dashed lines). Also shown is the distribution of measured separations for 20 $z \approx 10.6$ candidates⁴ in the HUDF (stepped blue histogram). Results in **a**, **b** and **c** correspond to samples of Lyman-break galaxy candidates selected with median redshifts of $z \approx 6$, $z \approx 7$, $z \approx 8.6$ and $z \approx 10.6$. At $z \approx 6$, candidate selection using the Advanced Camera for Surveys reaches⁶ $m_{lim} = 30$ mag (absolute magnitude $M_{lim} = -16.7$ mag). At higher redshifts, objects in the WFC3 HUDF data can be selected⁴ to $m_{lim} \approx 29.0$ mag, corresponding to $M_{lim} = -18.0$, -18.3 and -18.6 mag at $z \approx 7$, 8.6 and 10.6 . In **a** and **b** the open squares correspond to lens fractions given the fitting formula^{1,25} $M_* \approx -21 + 0.32 \times (z - 3.8)$, and the filled squares represent alternative estimates^{4,6} of M_* . The model for gravitational lensing²⁶ is based on the velocity dispersion function of galaxies²⁷. Galaxy mass distributions are modelled as singular isothermal spheres, and we assume a constant co-moving density of lenses. Elliptical lenses would not significantly alter the cross-section²⁸, but would provide additional images, and so increase the fraction of observed galaxies that are lensed. We assume a Schechter luminosity function²², with power-law slope¹ $\alpha = -2$. A change of 0.3 in α leads to a 40% change in the lens probability. We have used the cosmology based on 7-year results from the WMAP satellite²⁹ throughout this Letter.

distribution of redshifts for foreground galaxies found within $\Delta\theta < 1.5$ arcsec of the $z \approx 10.6$ candidates. These distributions are consistent with the distribution of gravitational lens redshifts, while the redshift distribution of all bright foreground galaxies are not, which supports the hypothesis that foreground galaxies are lensing a fraction of the $z \approx 10.6$ candidates into the observed sample.

With the introduction of the James Webb Space Telescope (JWST), galaxy surveys will be undertaken out to even higher redshifts, well into the epoch of first light²¹. We show F_{lens} as a function of M_* out to $z = 20$ in Fig. 2a, b. The flux limits correspond to an ultra-deep survey ($m_{lim} = 31.4$ mag), and a medium-deep survey ($m_{lim} = 29.4$ mag). The evolution of the characteristic luminosity is unknown at these unexplored redshifts. For comparison, we therefore plot squares corresponding to estimates of M_* based on an extrapolation from lower redshift HUDF data¹. Figure 2 shows that in ultra-deep JWST surveys for first light objects at $z \gtrsim 14$, more than $F_{lens} \approx 10\%$ of the candidates could be lensed. In much shallower JWST surveys that only sample the exponential tail of the Schechter luminosity function, a lensed object fraction of $F_{lens} \approx 10\%$ could be seen at redshifts as low as $z \approx 8$ – 10 . However at $z \gtrsim 14$, the lensed fraction in such surveys could be much higher, and may even represent the majority of observed galaxies. Surveys with JWST will therefore need to be carefully planned and analysed to account for the influence of foreground lensing galaxies.

As in the case of the HUDF, the fraction of galaxies that will be detected as multiply imaged systems by JWST is significantly lower than the true multiple image fraction. However, as the multiple image fraction becomes very large at high redshifts, observed doubles could become common; larger than $F_{mult} \approx 10\%$ at redshifts $z \gtrsim 12$ in a medium-deep ($m_{AB} < 29.4$ mag) JWST survey, and $z \gtrsim 16$ in an ultra-deep ($m_{AB} < 31.4$ mag) survey. In Fig. 2e, f, we present the predicted distributions of separation for galaxies discovered by JWST from bright foreground galaxies. If the observed evolution in M_* continues to higher redshift, then the spatial distribution of high-redshift galaxies relative to foreground galaxies will depart from random at redshifts $z \gtrsim 14$ for ultra-deep surveys, and at $z \gtrsim 10$ for medium-deep surveys with JWST. A crucial prediction is that the majority of very-high-redshift galaxies discovered with JWST may be located less than 1 arcsec from a bright foreground galaxy, and will have been gravitationally magnified into the sample.

A key goal for JWST will be to measure the number counts of high-redshift candidates, and to construct luminosity functions in order to build up a statistical description of star-forming activity in galaxies. Luminosity functions describing the density of sources per unit luminosity are parametrized by a Schechter function²², $\Psi(L) \propto (L/L_*)^\alpha \exp(-L/L_*)$, including free parameters for the power-law slope at low luminosities (α), and the characteristic absolute AB-magnitude ($M_{AB} - M_* = -2.5 \log_{10}(L/L_*)$) brighter than which galaxy numbers drop exponentially. Importantly, gravitational lensing has the potential to significantly modify the observed luminosity function from its intrinsic shape²³. In particular, at very high luminosities in the exponential tail of the Schechter function, the shape can be modified from exponential to power-law, since gravitational lensing magnifies numerous faint sources to apparently higher luminosities. Figure 3 shows that the shapes of luminosity functions near the flux limit are not affected by gravitational lensing at $z \approx 6$ – 8 . However, if the evolution of the galaxy luminosity function continues into the reionization era (we assume an extrapolation of the fitting formulae based on candidates discovered in and around the HUDF¹), then we find that JWST will measure luminosity functions that are significantly modified by lensing at redshifts above $z \approx 14$ and $z \approx 10$ in its ultra-deep and medium-deep surveys, respectively.

Our results imply that although published luminosity functions at $z \gtrsim 7$ are not currently corrected for a potential gravitational lensing bias, such corrections will need to be prescribed in detail for future surveys using JWST that aim to measure the build-up of stellar mass among the first galaxies. In particular, studies of the high-redshift

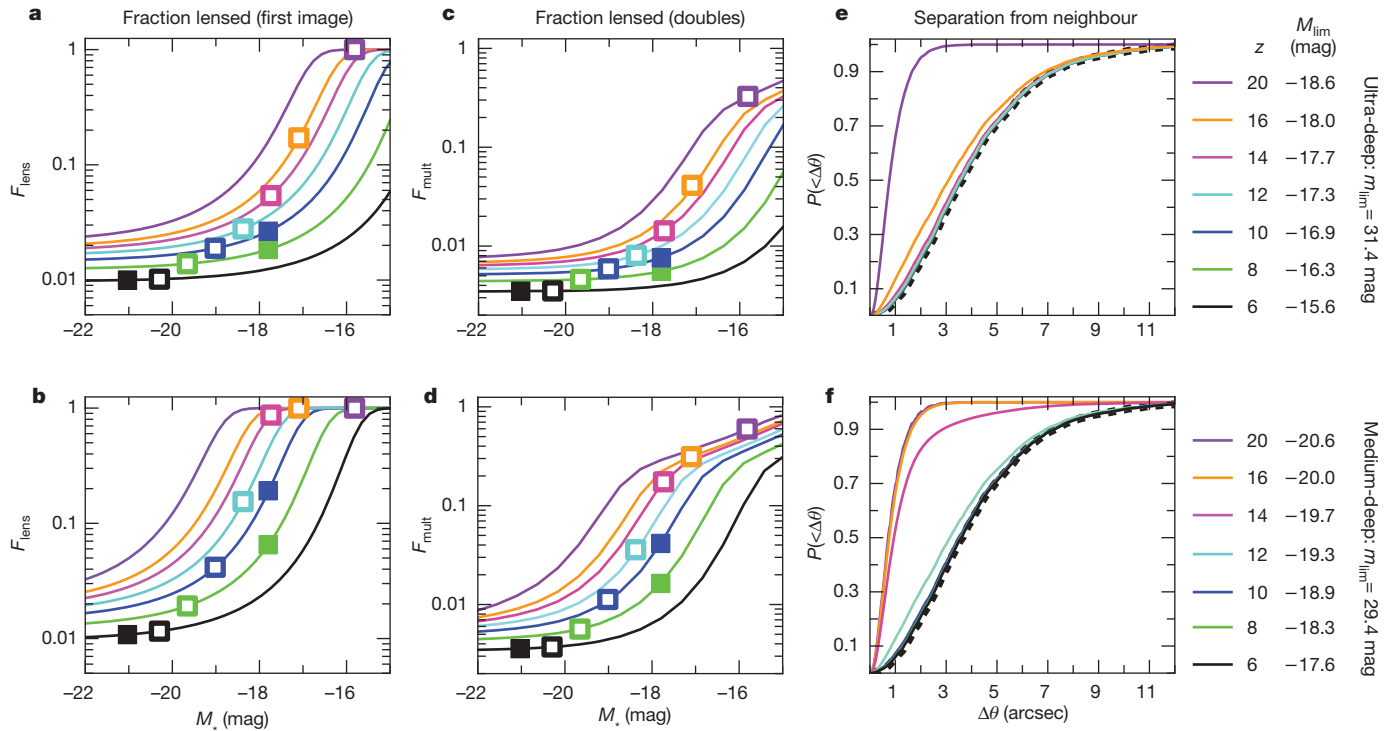


Figure 2 | Probabilities of multiple imaging of high-redshift galaxies with JWST. The panels mirror those of Fig. 1, but with examples of limiting magnitudes and redshifts appropriate for both an ultra-deep survey (top row: $m_{\text{lim}} = 31.4$ mag, ~ 1 nJy), and a medium-deep survey (bottom row: $m_{\text{lim}} = 29.4$ mag) with JWST. The corresponding limiting absolute magnitudes are listed at right. **a, b,** The fraction of observed galaxies that have multiple images (with $m_{\text{AB},1} < m_{\text{lim}}$). The superimposed filled and open points correspond to lens fractions given a faint value⁴ of M_* at $z \approx 8.6$ and $z \approx 10.6$, and a fitting formula $M_*(z)$ based on lower redshift data, respectively^{1,25}. The latter is extrapolated to high redshift where data does not yet exist. **c, d,** The fraction of high-redshift galaxies in which multiple images

could be detected by JWST (with $m_{\text{AB},2} < m_{\text{lim}}$). **e, f,** The probability distribution of image separations relative to the nearest bright foreground galaxy, in the cases of random lines of sight (black dotted line), and for composite distributions computed for values of M_* extrapolated from observations in the HUDF using the previously mentioned fitting formula^{1,25}. We note that imaging surveys with JWST will be working at the diffraction limit (~ 0.08 arcsec resolution FWHM) at $\sim 2 \mu\text{m}$. This resolution is higher than is currently available in the HUDF near-infrared images, where candidates have been selected in close proximity to bright foreground galaxies, and hence high-redshift candidates will also be detectable close to foreground galaxies.

luminosity function will require good understanding of the magnification bias for high-redshift galaxies, in order to correct for gravitational lensing and uncover its true unlensed shape at $z \gtrsim 12$. Of particular importance will be the unknown evolution of M_* , which could be influenced (for example) by supernova feedback from population-III stars²⁴, in addition to hierarchical clustering and formation. Gravitational lensing could magnify objects at $z \gtrsim 10$ –12 to flux levels that will allow spectroscopic observations using JWST and the largest ground-based near-infrared spectrographs. A further implication of our analysis is that gravitational lensing could be used to probe the shape of the high-redshift luminosity function at luminosities that are not otherwise accessible¹², using the association of high-redshift galaxy

candidates and foreground galaxies, combined with careful modelling of the gravitational lensing bias.

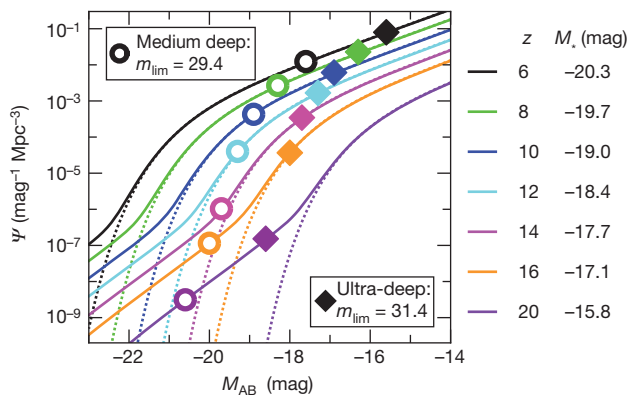


Figure 3 | Gravitational-lens-induced modification of the bright end of the high-redshift galaxy luminosity function to be observed with JWST. Dotted curves present the intrinsic luminosity function (ψ), and solid curves the observed luminosity function following modification from gravitational lensing. For simplicity, a uniform magnification was assumed outside regions of sky that are multiply imaged, with a value such that flux is conserved over the whole sky. The parameters describing the luminosity function are extrapolated to high redshift, where data do not yet exist, assuming fitting formulae based on data from the HUDF^{1,25}. Of particular relevance are the values of M_* , which are listed (right). The filled and open points show the luminosities and densities of the faintest galaxies to be observed with JWST, assuming limiting magnitudes appropriate for both an ultra-deep JWST survey ($m_{\text{AB}} < 31.4$ mag), and a medium-deep JWST survey ($m_{\text{AB}} < 29.4$ mag). The probability of gravitational lensing will become of order unity in the steep exponential parts of the luminosity function at sufficiently high redshifts. This gravitational forest should not be confused with the purely mathematical effects of image crowding that makes the detection and de-blending of faint objects harder at progressively fainter fluxes³⁰. These latter effects are referred to as either the instrumental confusion limit—when the instrumental resolution is not good enough to statistically distinguish all faint background objects from brighter foreground objects—or the natural confusion limit—when the instrumental resolution is good enough to distinguish faint background objects from brighter foreground objects, but the images are so deep that objects start overlapping because of their own intrinsic sizes. The HUDF and JWST images are in the latter regime³⁰, and as argued in this Letter, probably have the additional fundamental limitation that gravitational lensing will magnify a non-negligible fraction of faint objects into the sample.

Received 21 July; accepted 27 October 2010.

1. Bouwens, R. J. *et al.* UV luminosity functions from 113 $z \sim 7$ and $z \sim 8$ Lyman-break galaxies in the ultra-deep HUDF09 and wide-area ERS WFC3/IR observations. Preprint at (<http://arxiv.org/abs/1006.4360>) (2010).
2. Lorenzoni, S. *et al.* Candidate $z \sim 8-9$ galaxies from WFC3 imaging. Preprint at (<http://arxiv.org/abs/1006.3545>) (2010).
3. Bouwens, R. J. *et al.* Constraints on the first galaxies: $z \sim 10$ galaxy candidates from HST WFC3/IR. Preprint at (<http://arxiv.org/abs/0912.4263>) (2009).
4. Yan, H. *et al.* Galaxy formation in the reionization epoch as hinted by Wide Field Camera 3 observations of the Hubble Ultra Deep Field. *Res. Astron. Astrophys.* **10**, 867–904 (2010).
5. Bouwens, R. J. *et al.* Star formation at $z \sim 6$: the Hubble Ultra Deep Parallel Fields. *Astrophys. J.* **606**, L25–L28 (2004).
6. Yan, H. & Windhorst, R. A. Candidates of $z \sim 5.5-7$ galaxies in the Hubble Space Telescope Ultra Deep Field. *Astrophys. J.* **612**, L93–L96 (2004).
7. Wilkins, S. M. *et al.* Probing L^* Lyman-break galaxies at $z \sim 7$ in GOODS-South with WFC3 on Hubble Space Telescope. *Mon. Not. R. Astron. Soc.* **403**, 938–944 (2010).
8. McLure, R. J. *et al.* Galaxies at $z = 6-9$ from the WFC3/IR imaging of the Hubble Ultra Deep Field. *Mon. Not. R. Astron. Soc.* **403**, 960–983 (2010).
9. Yan, H. & Windhorst, R. A. The major sources of the cosmic reionizing background at $z \sim 6$. *Astrophys. J.* **600**, L1–L5 (2004).
10. Trenti, M. *et al.* The galaxy luminosity function during the reionization epoch. *Astrophys. J.* **714**, L202–L207 (2010).
11. Barkana, R. & Loeb, A. High-redshift galaxies: their predicted size and surface brightness distributions and their gravitational lensing probability. *Astrophys. J.* **531**, 613–623 (2000).
12. Comerford, J. M., Haiman, Z. & Schaye, J. Constraining the redshift $z \sim 6$ quasar luminosity function using gravitational lensing. *Astrophys. J.* **580**, 63–72 (2002).
13. Turner, E. L., Ostriker, J. P. & Gott, J. R. III. The statistics of gravitational lenses—the distributions of image angular separations and lens redshifts. *Astrophys. J.* **284**, 1–22 (1984).
14. Richard, J. *et al.* A Hubble and Spitzer Space Telescope survey for gravitationally lensed galaxies: further evidence for a significant population of low-luminosity galaxies beyond $z = 7$. *Astrophys. J.* **685**, 705–724 (2008).
15. Bradley, L. D. *et al.* Discovery of a very bright strongly lensed galaxy candidate at $z \sim 7.6$. *Astrophys. J.* **678**, 647–654 (2008).
16. Zheng, W. *et al.* Bright strongly lensed galaxies at redshift $z \sim 6-7$ behind the clusters Abell 1703 and CL0024+16. *Astrophys. J.* **697**, 1907–1917 (2009).
17. Bouwens, R. J. *et al.* $z \sim 7-10$ galaxies behind lensing clusters: contrast with field search results. *Astrophys. J.* **690**, 1764–1771 (2009).
18. Webster, R. L., Hewett, P. C., Harding, M. E. & Wegner, G. A. Detection of statistical gravitational lensing by foreground mass distributions. *Nature* **336**, 358–359 (1988).
19. Nollenberg, J. G. & Williams, L. L. R. Galaxy-quasar correlations between APM galaxies and Hamburg-ESO QSOs. *Astrophys. J.* **634**, 793–805 (2005).
20. Scranton, R. *et al.* Detection of cosmic magnification with the Sloan Digital Sky Survey. *Astrophys. J.* **633**, 589–602 (2005).
21. Windhorst, R. A., Cohen, S. H., Jansen, R. A., Conselice, C. & Yan, H. How JWST can measure first light, reionization and galaxy assembly. *N. Astron. Rev.* **50**, 113–120 (2006).
22. Schechter, P. An analytic expression for the luminosity function for galaxies. *Astrophys. J.* **203**, 297–306 (1976).
23. Pei, Y. C. Magnification of quasars by cosmologically distributed gravitational lenses. *Astrophys. J.* **440**, 485–500 (1995).
24. Greif, T. H., Johnson, J. L., Bromm, V. & Klessen, R. S. The first supernova explosions: energetics, feedback, and chemical enrichment. *Astrophys. J.* **670**, 1–14 (2007).
25. Hathi, N. P. *et al.* UV-dropout galaxies in the GOODS-South Field from WFC3 early release science observations. *Astrophys. J.* **720**, 1708–1716 (2010).
26. Oguri, M. *et al.* The Sloan Digital Sky Survey quasar lens search. III. Constraints on dark energy from the third data release quasar lens catalog. *Astron. J.* **135**, 512–519 (2008).
27. Choi, Y., Park, C. & Vogeley, M. S. Internal and collective properties of galaxies in the Sloan Digital Sky Survey. *Astrophys. J.* **658**, 884–897 (2007).
28. Huterer, D., Keeton, C. R. & Ma, C. Effects of ellipticity and shear on gravitational lens statistics. *Astrophys. J.* **624**, 34–45 (2005).
29. Komatsu, E. *et al.* Five-year Wilkinson Microwave Anisotropy Probe observations: cosmological interpretation. *Astrophys. J., Suppl.* **180**, 330–376 (2009).
30. Windhorst, R. A. *et al.* High resolution science with high redshift galaxies. *Adv. Space Res.* **41**, 1965–1971 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K.-H. Chae for discussing results of his lensing calculations. J.S.B.W. was supported in part by a QE-II fellowship and grants from the Australian Research Council. H.Y. acknowledges support from the long-term fellow-ship programme of the Center for Cosmology and AstroParticle Physics (CCAPP) at The Ohio State University. H.Y. and R.A.W. were supported by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc. R.A.W. was supported by a NASA JWST Interdisciplinary Scientist grant.

Author Contributions J.S.B.W. performed the calculations of lensing probabilities. H.Y. measured the distributions of galaxy properties from the HUDF. All authors were involved in the conception of the work, discussing the results, and writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S.B.W. (swyithe@unimelb.edu.au).

Local charge of the $\nu = 5/2$ fractional quantum Hall state

Vivek Venkatachalam¹, Amir Yacoby¹, Loren Pfeiffer² & Ken West²

Electrons moving in two dimensions under the influence of strong magnetic fields effectively lose their kinetic energy and display exotic behaviour dominated by Coulomb forces. When the ratio of electrons to magnetic flux quanta in the system (ν) is near $5/2$, the electrons are predicted to condense into a correlated phase with fractionally charged quasiparticles and a ground-state degeneracy that grows exponentially as these quasiparticles are introduced¹. The only way for electrons to transform between the many ground states would be to braid the fractional excitations around each other. This property has been proposed as the basis of a fault-tolerant quantum computer². Here we present observations of localized quasiparticles at $\nu = 5/2$, confined to puddles by disorder. Using a local electrometer to compare how quasiparticles at $\nu = 5/2$ and $\nu = 7/3$ charge these puddles, we were able to extract the ratio of local charges for these states. Averaged over several disorder configurations and samples, we found the ratio to be $4/3$, suggesting that the local charges are $e_{7/3}^* = e/3$ and $e_{5/2}^* = e/4$, where e is the charge of an electron. This is in agreement with theoretical predictions for a paired state at $\nu = 5/2$. Confirming the existence of localized $e/4$ quasiparticles shows that proposed interferometry experiments to test statistics and computational ability of the state at $\nu = 5/2$ would be possible.

When a two-dimensional electron system (2DES) is subject to a strong perpendicular magnetic field, the physics that emerges is controlled by interelectron Coulomb interactions. If the 2DES is tuned such that ν is near certain rational values, the electrons condense into so-called fractional quantum Hall (FQH) phases³. These strongly correlated states are gapped and incompressible in the bulk of the sample, but metallic and compressible along the sample boundary, allowing current to flow around the perimeter in such a way that the transverse conductance (G_{xy}) is precisely quantized to $G_{xy} = \nu (e^2/h)$, where e is the electron charge and h is Planck's constant. Additionally, the electronic correlations encoded in FQH states give rise to local excitations with a charge a fraction of that of an electron, and braiding statistics that fall outside the conventional classification of bosonic or fermionic. It is predicted that particle interchange can adiabatically evolve a $\nu = 5/2$ system between orthogonal ground states, a property not seen in the state's conventional odd-denominator relatives¹. This property, dubbed non-Abelian braiding statistics, has been proposed as the basis for a topological quantum computer that would be insensitive to environmental decoherence^{2,4}.

One necessary (but insufficient) condition for the existence of exotic braiding statistics at $\nu = 5/2$ is for the ground state to support local excitations with a charge of $e_{5/2}^* = e/4$ (ref. 1). Although a charge of $e/4$ had previously been measured using shot-noise techniques⁵, more-recent data from the same group⁶ suggest that the value of the measured charge changes continuously as the point-contact conductance and temperature are varied, reaching an inferred charge of unity in the weak and strong tunnelling limits. Unexpected charges have also been reported for the more conventional fractions at $1/3$, $2/3$ and $7/3$ (refs 7, 6). Moreover, conductance measurements in the weak tunnelling regime⁸ suggest a quasiparticle charge of $e_{5/2}^* = 0.17e$, in stark contrast to the shot-noise results.

Clearly, a better understanding of the tunnelling processes that take place between quantum Hall edges in the quantum point contact is needed to interpret the shot-noise results. Alternatively, one can infer quasiparticle charge using a thermodynamic approach⁹ that probes the quasiparticle charge in the bulk of the sample. Here we use a single-electron transistor (SET) as a sensitive electrometer to measure the equilibrium charge distribution in the bulk and determine its dependence on the average density and magnetic field. Our results provide clear evidence for quasiparticles at $\nu = 5/2$ with a localized charge of $e/4$.

Our measurement used a fixed SET as a gated device capable of sensitively measuring the local incompressibility ($\kappa^{-1} = \frac{\partial \mu}{\partial n}$, where μ is the local chemical potential and n is the global electron density) of a high-mobility 2DES (ref. 10). The 2DES was in a GaAs/AlGaAs quantum well, grown by molecular beam epitaxy, that was 200 nm deep and 30 nm wide, with symmetric silicon δ -doping layers 100 nm on either side. A metallic back-gate grown 2 μm below the 2DES allowed us to tune n in the well over a typical range of $(2.3\text{--}2.5) \times 10^{15} \text{ m}^{-2}$, with some variation between samples. The SET was fabricated on top of the sample using standard electron-beam lithography and shadow-evaporation techniques (Fig. 1), creating an island

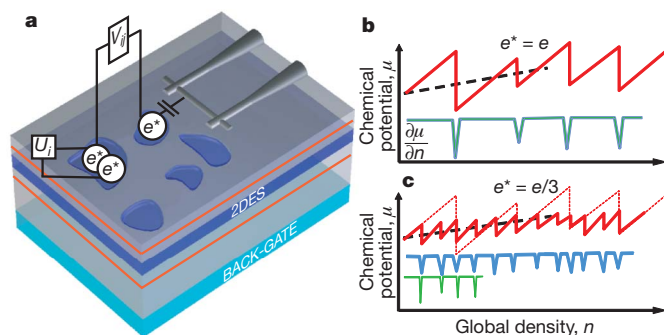


Figure 1 | Filling puddles with fractional charge. **a**, Electrometer used to measure the equilibrium charge distribution in the bulk sample. The quantum well (dark blue, marked 2DES) is 30 nm wide, with symmetric Si δ -doping layers 100 nm on either side (orange bands). Donors in these layers create a disorder potential in the 2DES, which produce puddles of localized states when the bulk is tuned to an incompressible, percolating Hall state. These puddles have some charging energy associated with adding electrons (U_i), and possibly some interaction with surrounding puddles (V_{ij}). Incompressibility ($\kappa^{-1} = \frac{\partial \mu}{\partial n}$) is measured using an SET fabricated on the surface. **b**, Whereas the global chemical potential should increase smoothly with density (black dashed line), the local chemical potential will increase in jumps (red line), with charge being added when the global chemical potential aligns with a localized state. The green line shows the incompressibility of the sample. **c**, Repeating the charging of an identical puddle with objects of charge $e/3$ (solid red line) rather than charge e (dashed red line) results in three times as many charging events in the same range of global density. Scaling the density axis of the charge e spectrum by $1/3$ and shifting it by some amount (green line) should result in good overlap of the incompressibility spectra (blue and green lines).

¹Department of Physics, Harvard University, 11 Oxford Street, Cambridge, Massachusetts 02138, USA. ²Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA.

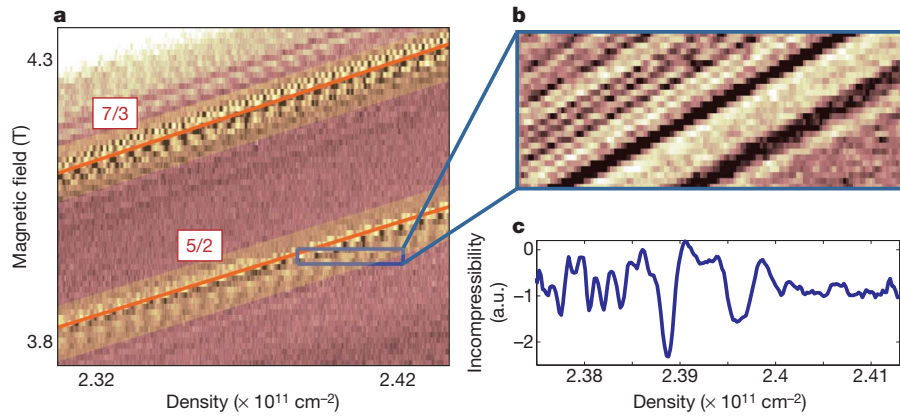


Figure 2 | Incompressibility and localized states at 5/2. **a**, By varying the magnetic field and the back-gate voltage (density), we can identify incompressible phases of the 2DES. Our samples show clear incompressible FQH states at 5/2 and 7/3, with the expected slopes in the nB -plane. **b**, Zooming in shows repeatable

with dimensions $500 \text{ nm} \times 80 \text{ nm}$. All measurements were carried out in a dilution refrigerator with an electron temperature of 20 mK, verified using standard Coulomb blockade techniques.

As we adjust the density and magnetic field (B) we expect to see regions of incompressibility when a gap is present, which will only happen precisely when the system is in a quantum Hall state. The slope of these incompressible regions in the nB -plane corresponds to the filling factor of the state¹¹. Figure 2 shows incompressibility versus density and magnetic field between $\nu = 2$ and $\nu = 3$, with the two highlighted regions corresponding to FQH states at $\nu = 5/2$ and $\nu = 7/3$.

Additionally, owing to the rough disorder potential created by remote donors, we expect different points in space to develop gaps at different values of n . Because of this, we expect a well-developed quantum Hall state to have a percolating incompressible region punctured by small compressible puddles that behave as either dots or antidots¹¹. As n is varied, a given compressible puddle will occasionally be populated by quasiparticles or quasiholes of the surrounding incompressible state. This creates a jump in the local chemical potential, $\mu(n)$, and a spike in the local incompressibility ($\frac{\partial \mu}{\partial n}$). The magnitude and spacing of these spikes is determined by the charging spectrum of the puddle, which in turn is dictated by the quasiparticle charge in the surrounding incompressible region. That is, if the quasiparticle charge were reduced by a factor of three for a fixed disorder potential, we should see three times as many compressible spikes in μ as a function of n (Fig. 1b,c).

This difference in spike frequencies has been used to measure the local charge at $\nu = 1/3$ and $\nu = 2/3$ (ref. 9). Unlike shot-noise measurements⁷, these local compressibility measurements find a quasiparticle charge of $e/3$ at both filling factors. Furthermore, because of the spatial resolution afforded by the scanning technique in that measurement, it was possible to establish that the disorder potential landscape does not change as the electron system is tuned between Hall states with comparable gaps. Transport measurements confirm that the gap inferred from activation of minima in the longitudinal resistance (R_{xx}) is comparable for the states at 5/2 and 7/3 (refs 12, 13), so we can expect similar potential landscapes for the two states.

Our procedure began with obtaining charging spectra (incompressibility versus density) at $\nu = 5/2$ and $\nu = 7/3$. Because the gap for these states is comparable, and the disorder potential is not altered as we change the magnetic field or density, we expect the spacing between charging features to reflect the quasiparticle charge in each state. In the limit of an isolated compressible puddle surrounded by an incompressible fluid, this relationship is particularly simple: if the ratio of local charges between the two spectra is β , the spectra should be identical after one of the density axes is rescaled by a factor of β , and shifted by

charging events associated with quasiparticles localizing in puddles under the SET, stable on a timescale of days. **c**, A linecut showing the charging spectrum of any puddles coupled to the SET. Downwards spikes correspond to quasiparticles entering puddles beneath the SET. a.u., arbitrary units.

some amount (Fig. 3a). To proceed, we chose a value of β and stretched one of the spectra by this factor. We then calculated the correlation $\left(\frac{\langle C_1(x)C_2(x) \rangle}{\sqrt{\langle C_1(x)^2 \rangle \langle C_2(x)^2 \rangle}} \right)$ between the two spectra as a function of density offset and recorded the highest value. We repeated this process for many scaling factors to obtain the best correlation versus β , as depicted in Fig. 3b.

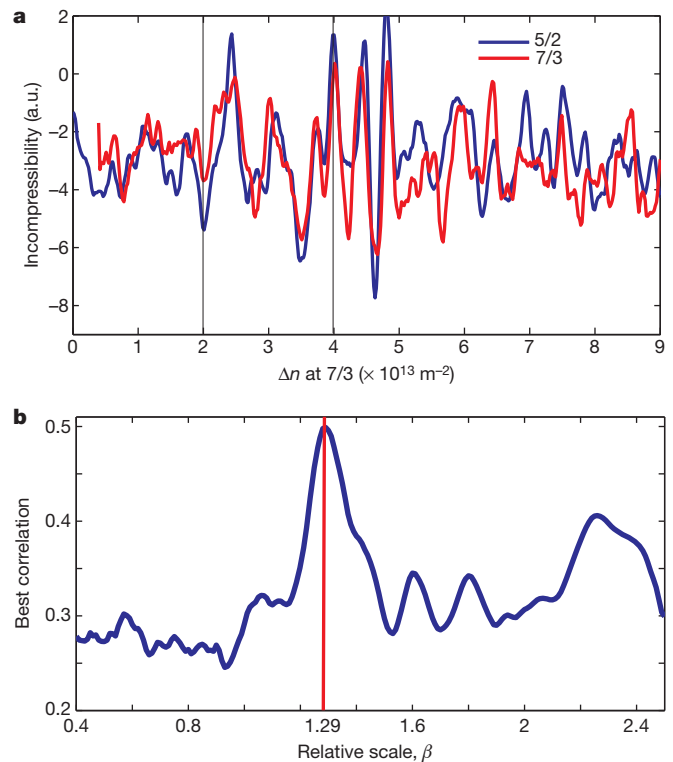


Figure 3 | Comparison of spectra at 5/2 and 7/3. **a**, To determine the charge, we first chose a relative scale between the two density axes (β) and determined the offset between the two spectra that maximized the cross-covariance. Here, the density for the spectrum at 5/2 is scaled up by a factor of 1.29 and shifted to match the spectrum at 7/3. The guide lines show the density change required to add one electron to an area of $100 \text{ nm} \times 500 \text{ nm}$, roughly the size of our SET. We would therefore expect, very roughly, $3 e/3$ charging events in a window this size. Δn , variation in density for the 7/3 (unscaled) trace. **b**, Repeating this for many values of β suggests that a relative scale of 1.29 best describes this data set.

This procedure was repeated for 20 different disorder configurations, obtained by changing samples, measuring with different SETs or thermal cycling to change the disorder. A summary of the data is shown in Fig. 4a, with an average over the measured ensemble in Fig. 4b. Further details regarding the analysis procedure can be found in the Supplementary Material. The peak observed at $\beta = 1.31$ suggests a charge ratio of 4/3 between the two states, and a qualitative inspection of spectra overlap (as in Fig. 3a) corroborates this. To determine the significance of the peak value, we repeated our analysis with pairs of spectra from different disorder configurations, which should be less well correlated. For each scale, we characterized the distribution of best correlations with a mean and standard deviation. These, in turn, can be simply converted to the expected mean and standard error for our data (if it were uncorrelated). The 1σ region around the uncorrelated mean is depicted in red in Fig. 3b. Our averaged correlation at $\beta = 1.31$ lies 3.8 standard errors above the uncorrelated mean, corresponding to a one-tailed P -value of 7×10^{-5} . Assuming a charge of $e_{7/3}^* = e/3$, this measured value of β suggests $e_{5/2}^* = (e/3)/(1.31) = 0.254e$, in agreement with the Moore–Read prediction of $e_{5/2}^* = e/4$ (ref. 1).

To better understand why some configurations seem to provide weaker (and sometimes different) measurements of β , it helps to abandon the assumption that we are charging and monitoring single puddles, as well as the assumption that quasiparticles in different puddles do not interact. A free energy for our system that takes these factors into account is given by:

$$F = \sum_i (\varepsilon_i - V_{BG}) Q_i + \frac{1}{2} \sum_i U_i Q_i (Q_i - 1) + \sum_{i < j} V_{ij} Q_i Q_j - \sum_i \Delta \left[\frac{Q_i}{2} \right]$$

Here, U_i and ε_i are the on-site interaction (self-capacitance) and bare disorder potential for puddle i , respectively. V_{ij} is a pairwise interaction, or cross-capacitance, between puddles i and j , Q_i and Q_j are the charges for puddles i and j respectively, and Δ is the energy gained by forming a bound pair of quasiparticles. V_{BG} is the voltage applied to the backgate and is proportional to the overall number of particles

added. For now, we will let $\Delta = 0$. We assumed that some subset of the puddles is capacitively coupled to and measured by the SET.

To compute charging spectra from this model, we first chose values of U , V and ε for each puddle from Gaussian distributions. We then divided Q_i into discrete units of $e/3$ or $e/4$ and determined how many units of charge to put in each puddle to minimize the above free energy. This was done for each value of V_{BG} , and converted into a charging spectrum. Finally, we took the resulting spectra and repeated the processing performed on the data to obtain summary statistics for comparison. The result, with $\varepsilon = 0 \pm 0.3U$ and $V_{ij} = 0.3U \pm 0.2U$, is shown in Fig. 4c. Results for other parameter choices in a large range are qualitatively similar, with smaller values of σ_ε (uncertainty in ε) and V_{ij} corresponding to sharper peaks and less spread. As expected, these simulations tell us that both ε and V_{ij} can distort spectra in such a way that the maximum cross-covariance will shift slightly or even dramatically away from 4/3. Still, we should always expect some weight at 4/3, and this can be extracted by averaging over disorder configurations (Fig. 4d). There exist two further factors that may contribute to irregular peak spacing in our data that are not considered in the model. First, the 2DES is only truly incompressible for a small range of density, as it transitions into and out of the FQH phase. Outside this range, there is further screening from compressible regions near the puddles of interest that can reduce the effect of the back-gate. Second, there may be quantum effects beyond electrostatics that may alter how quasiparticles in these states charge small puddles.

Recently, there has been some suggestion that $e/2$ quasiparticles may be present at the $\nu = 5/2$ edge and may be relevant to interference measurements¹⁴. In the context of our model, we can consider the effect that weak binding of quasiparticles would have on measured spectra. This binding can be described in terms of the parameter Δ above, and we consider only the case in which pairing affects the $e/4$ quasiparticles. As the strength of pairing is increased relative to the on-site interaction (Fig. 4d), we expect weight to shift from the peak at 4/3 to a peak at 2/3 (corresponding to $e/2$ quasiparticles), with considerable weight at 2/3 even when $\Delta = 0.1U$. Our data show no appreciable evidence for a peak at 2/3, suggesting that the only quasiparticles participating in localization have charge $e/4$. This, however, does not exclude the possibility of $e/2$ excitations on the edge.

To our knowledge, these measurements constitute the first direct measurement of incompressibility and localized states at $\nu = 5/2$, and they provide an equilibrium probe of the local charge that is insensitive to complications arising from measurements of transport through nanostructures. The measured value, $e_{5/2}^* = e/4$, indicates that the FQH state at $\nu = 5/2$ demonstrates pairing, in agreement with proposed non-Abelian variational wavefunctions and differing from other observed FQH states. Finally, the localization of $e/4$ quasiparticles is essential to the development of interferometers capable of detecting and exploiting these exotic braiding properties^{15,16}, and our measurements suggest that $e/4$ localization does indeed occur in a well-behaved way.

Received 14 May; accepted 16 November 2010.

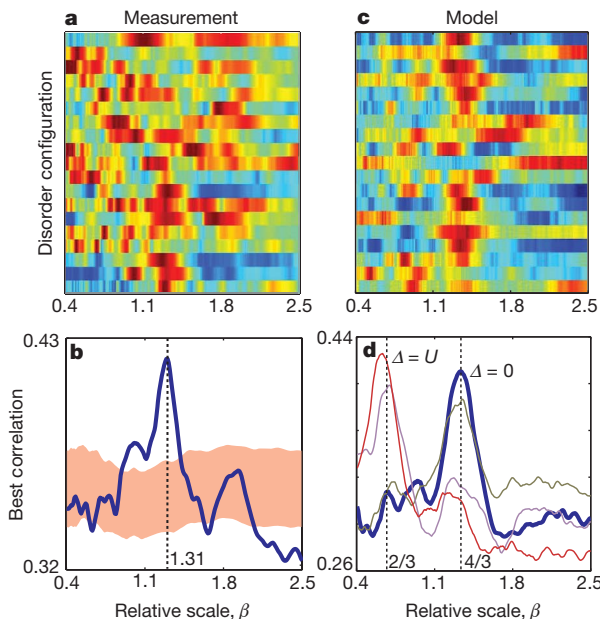


Figure 4 | Summary of data and model. **a**, Repeating the measurement over many disorder configurations and samples shows that the peak at 4/3 is usually present. **b**, Averaging over all measurements yields a clear peak at $\beta = 1.31$, 3.8σ ($P = 7 \times 10^{-5}$) above the uncorrelated background for that scale, colored in orange, suggesting a local charge ratio of 4/3. **c**, **d**, Running our model with parameters $\varepsilon = 0 \pm 0.3$, $V = 0.3 \pm 0.2$ and $\Delta_{5/2} = 0$ (blue), 0.01 (grey), 0.1 (purple) and 1.0 (red) (all in units of U , the on-site charging energy). We simulated charging of four puddles, of which two were capacitively coupled to the SET.

- Moore, G. & Read, N. Nonabelions in the fractional quantum hall effect. *Nucl. Phys. B* **360**, 362–396 (1991).
- Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. Non-abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
- Girvin, S. & Prange, R. *The Quantum Hall Effect* (Springer, 1987).
- Das Sarma, S., Freedman, M. & Nayak, C. Topologically protected qubits from a possible non-abelian fractional quantum hall state. *Phys. Rev. Lett.* **94**, 166802 (2005).
- Dolev, M., Heiblum, M., Umansky, V., Stern, A. & Mahalu, D. Observation of a quarter of an electron charge at the $\nu = 5/2$ quantum Hall state. *Nature* **452**, 829–834 (2008).
- Dolev, M. *et al.* Dependence of the tunneling quasiparticle charge determined via shot noise measurements on the tunneling barrier and energetics. *Phys. Rev. B* **81**, 161303 (2010).
- Bid, A., Ofek, N., Heiblum, M., Umansky, V. & Mahalu, D. Shot noise and charge at the 2/3 composite fractional quantum Hall state. *Phys. Rev. Lett.* **103**, 236802 (2009).
- Radu, I. *et al.* Quasi-particle properties from tunneling in the $\nu = 5/2$ fractional quantum hall state. *Science* **320**, 899–902 (2008).

9. Martin, J. *et al.* Localization of fractionally charged quasi-particles. *Science* **305**, 980–983 (2004).
10. Ilani, S., Yacoby, A., Mahalu, D. & Shtrikman, H. Microscopic structure of the metal–insulator transition in two dimensions. *Science* **292**, 1354–1357 (2001).
11. Ilani, S. *et al.* The microscopic nature of localization in the quantum hall effect. *Nature* **427**, 328 (2004).
12. Dean, C. R. *et al.* Intrinsic gap of the $\nu = 5/2$ fractional quantum Hall state. *Phys. Rev. Lett.* **100**, 146803 (2008).
13. Choi, H. C., Kang, W., Das Sarma, S., Pfeiffer, L. N. & West, K. W. Activation gaps of fractional quantum hall effect in the second landau level. *Phys. Rev. B* **77**, 081301 (2008).
14. Bishara, W., Bonderson, P., Nayak, C., Shtengel, K. & Slingerland, J. K. Interferometric signature of non-abelian anyons. *Phys. Rev. B* **80**, 155303 (2009).
15. Stern, A. & Halperin, B. I. Proposed experiments to probe the non-abelian $\nu = 5/2$ quantum Hall state. *Phys. Rev. Lett.* **96**, 016802 (2006).
16. Bonderson, P., Kitaev, A. & Shtengel, K. Detecting non-abelian statistics in the $\nu = 5/2$ fractional quantum Hall state. *Phys. Rev. Lett.* **96**, 016803 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to acknowledge B. Verdene, J. Waissman and J. Nübler for technical assistance, and B. Halperin for theoretical discussions. We acknowledge financial support from Microsoft Corporation Project Q and the National Science Foundation's Graduate Research Fellowship Program.

Author Contributions V.V. conceived and designed the experiments, prepared samples, carried out the experiments and data analysis, and wrote the paper. A.Y. conceived and designed the experiments, carried out data analysis and wrote the paper. L.P. and K.W. carried out the molecular beam epitaxy growth.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.Y. (yacoby@physics.harvard.edu).

Two-dimensional electron gas with universal subbands at the surface of SrTiO₃

A. F. Santander-Syro^{1,2}, O. Copie^{3,4,5}, T. Kondo⁶, F. Fortuna¹, S. Pailhès⁷, R. Weht^{8,9}, X. G. Qiu¹⁰, F. Bertran¹¹, A. Nicolaou¹¹, A. Taleb-Ibrahimi¹¹, P. Le Fèvre¹¹, G. Herranz¹², M. Bibes^{3,4}, N. Reyren^{3,4}, Y. Apertet¹³, P. Lecoeur¹³, A. Barthélémy^{3,4} & M. J. Rozenberg^{14,15}

As silicon is the basis of conventional electronics, so strontium titanate (SrTiO₃) is the foundation of the emerging field of oxide electronics^{1,2}. SrTiO₃ is the preferred template for the creation of exotic, two-dimensional (2D) phases of electron matter at oxide interfaces^{3–5} that have metal–insulator transitions^{6,7}, superconductivity^{8,9} or large negative magnetoresistance¹⁰. However, the physical nature of the electronic structure underlying these 2D electron gases (2DEGs), which is crucial to understanding their remarkable properties^{11,12}, remains elusive. Here we show, using angle-resolved photoemission spectroscopy, that there is a highly metallic universal 2DEG at the vacuum-cleaved surface of SrTiO₃ (including the non-doped insulating material) independently of bulk carrier densities over more than seven decades. This 2DEG is confined within a region of about five unit cells and has a sheet carrier density of ~ 0.33 electrons per square lattice parameter. The electronic structure consists of multiple subbands of heavy and light electrons. The similarity of this 2DEG to those reported in SrTiO₃-based heterostructures^{6,8,13} and field-effect transistors^{9,14} suggests that different forms of electron confinement at the surface of SrTiO₃ lead to essentially the same 2DEG. Our discovery provides a model system for the study of the electronic structure of 2DEGs in SrTiO₃-based devices and a novel means of generating 2DEGs at the surfaces of transition-metal oxides.

In the cubic phase, stoichiometric SrTiO₃ has an empty t_{2g} conduction manifold composed of three dispersive, three-dimensional (3D) bands that are degenerate at the Γ point¹⁵. As schematically shown in Fig. 1a, its band structure along one direction, say k_y , consists of a weakly dispersive (heavy-mass) band and a pair of degenerate, strongly dispersive (light-mass) bands. They arise respectively from the small and large overlaps, along the y axis, of neighbouring titanium $3d_{xz}$, $3d_{xy}$ and $3d_{yz}$ orbitals, as depicted in Fig. 1b for the $3d_{xy}$ orbitals. Thus, the d_{xy} -like band will be light in the x - y plane and heavy along the z axis, whereas the d_{yz} band will be heavy along x and light in the y - z plane, and the d_{xz} band will be heavy along y and light in the x - z plane. The resulting three Fermi surface sheets, projected on the x - y plane of the sample surface, are shown in Fig. 1c for several bulk dopings.

Angle-resolved photoemission spectroscopy (ARPES) is a powerful technique for probing the electronic structure of materials. Previous work using ARPES addressed the bulk electronic structure of doped SrTiO₃, revealing dispersing quasiparticle peaks and in-gap features, and discussing possible polaronic effects near the Fermi energy, E_F (refs 16–19). As we shall see, our ARPES data do not conform to the expected lightly doped bulk band picture, but reveal that a novel 2DEG with a complex subband structure, shown schematically in Fig. 1e, is realized at the low-temperature vacuum-cleaved surface of this material.

Crucial to our discoveries is the comparison of samples with very different bulk carrier densities of $n_{3D} < 10^{13} \text{ cm}^{-3}$ (essentially non-doped), $n_{3D} \approx 10^{18} \text{ cm}^{-3}$ (low doping) and $n_{3D} \approx 10^{20} \text{ cm}^{-3}$ (high doping), determined by bulk-sensitive techniques (Supplementary Information, section II). The large difference in doping is apparent even by inspection of the samples (Fig. 2a–c): whereas the non-doped sample is transparent, the highly doped sample is black and shiny.

Our first observation (Fig. 2d) is that even for the transparent, non-doped sample (Fig. 2a), for which no bulk bands at E_F are expected, the cleaved surface yields intense, strongly dispersive bands across the Fermi level. They correspond to a large density of mobile carriers (see below), ensuring that, despite the electron emission, there is no charging of the surface of this otherwise highly insulating sample in the bulk. Equally notable is the fact (Fig. 2d–f) that for all the bulk dopings studied, the electronic states around E_F , and in particular the bandwidths (E_F minus the band bottom) and Fermi momenta, k_F , of the observed bands, are essentially identical. All these observations contrast sharply with the large differences that would be expected from electron-doping in the bulk (Fig. 1c), where the 3D density of carriers scales as k_F^3 . The immediate conclusion is that the observed bands correspond to 2D states (see also Supplementary Information, section III).

The spectra in Fig. 2d–f, measured around Γ_{102} and Γ_{012} (points such as B in Fig. 2j), clearly show two strongly dispersing bands, termed upper and lower parabolic bands. In Fig. 2g–i, two other weakly dispersing bands, termed upper and lower shallow bands, are revealed either by rotating the light polarization to be at right angles with respect to the measurement direction, as illustrated for the non-doped sample in Fig. 2g, or by measuring around a Γ point at a larger emission angle in a different Brillouin zone (Γ_{112} , or point C in Fig. 2j), as shown for the low-doped sample in Fig. 2h, i. We note furthermore, from Fig. 2d, g, that the changes in light polarization affect the two strongly dispersing parabolic bands differently, completely suppressing the upper parabolic band in Fig. 2g, whereas the increase in emission angle (compare Fig. 2e with Fig. 2h, i) enhances in the same asymmetric way the negative- k intensities of the two shallow bands and weakens the parabolic bands. These dichroic effects, which are due to dipole transition selection rules¹⁷, unambiguously show that the two parabolic bands have different symmetries, which are also different from a third symmetry common to the two shallow subbands (see also Supplementary Information, section IV). This means that the observed 2D bands are organized by their orbital symmetries.

Figure 3 illustrates the complete set of four bands (dashed lines) overlaid on the data, with the colours of the traced bands representing

¹CSNSM, CNRS/IN2P3 and Université Paris-Sud, Bâtiments 104 et 108, 91405 Orsay cedex, France. ²Laboratoire Physique et Etude des Matériaux, UMR 8213 ESPCI-CNRS-UPMC, 10 rue Vauquelin, 75231 Paris cedex 5, France. ³Unité Mixte de Physique CNRS/Thales, 1 Avenue A. Fresnel, Campus de l'Ecole Polytechnique, 91767 Palaiseau, France. ⁴Université Paris-Sud, 91405 Orsay, France. ⁵Universität Würzburg, Experimentelle Physik VII, 97074 Würzburg, Germany. ⁶Ames Laboratory and Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA. ⁷Laboratoire Léon Brillouin, CEA-CNRS, CEA-Saclay, 91191 Gif-sur-Yvette, France. ⁸Gerencia de Investigación y Aplicaciones, Comisión Nacional de Energía Atómica, Avenida General Paz y Constituyentes, 1650 San Martín, Argentina. ⁹Instituto Sábato, Universidad Nacional de San Martín – CNEA, 1650 San Martín, Argentina. ¹⁰Institute of Physics and National Laboratory for Condensed Matter Physics, Chinese Academy of Sciences, Zhongguancun nansanjie 8, Beijing 100190, China. ¹¹Synchrotron SOLEIL, CNRS-CEA, L'Orme des Merisiers, Saint-Aubin-BP48, 91192 Gif-sur-Yvette, France. ¹²Institut de Ciència de Materials de Barcelona, ICMA-B-CSIC, Campus de la UAB, 08193 Bellaterra, Catalonia, Spain. ¹³Institut d'Electronique Fondamentale, Université Paris-Sud, Bâtiment 220, 91405 Orsay, France. ¹⁴Laboratoire de Physique des Solides, Université Paris-Sud, Bâtiment 510, 91405 Orsay, France. ¹⁵Departamento de Física, FCEN-UBA, Ciudad Universitaria, Pabellón 1, Buenos Aires (1428), Argentina.

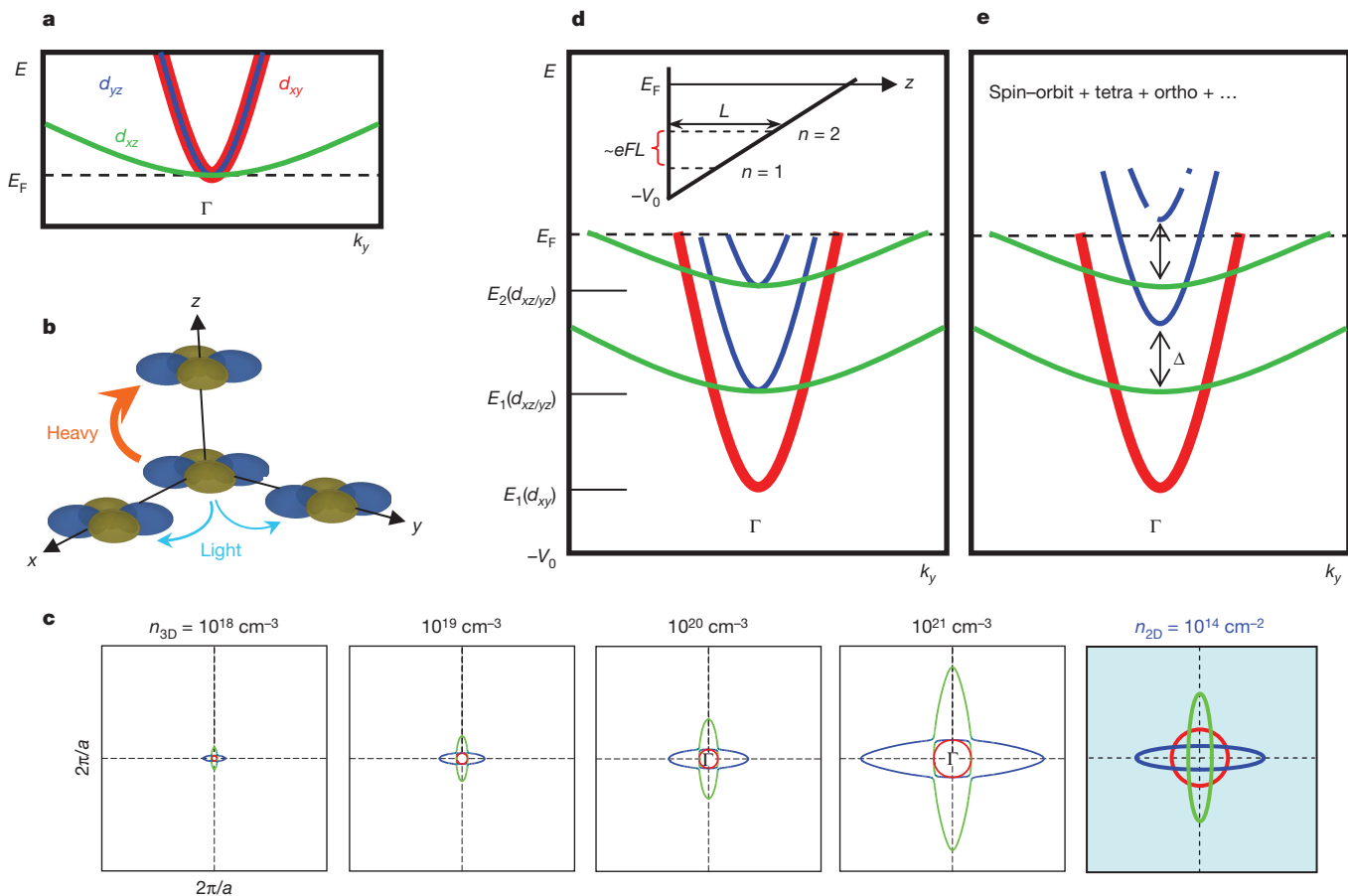


Figure 1 | Electronic structure of SrTiO₃ and effects of electron confinement. **a**, Bulk conduction band of SrTiO₃ along k_y , consisting of a heavy d_{xz} band (green) and a doublet of light d_{xy}/d_{yz} bands (red and blue). **b**, The bands in **a** stem from the small and large overlaps of the titanium 3d orbitals along the y direction, depicted here for the case of $3d_{xy}$ orbitals. **c**, Resulting 3D Fermi surfaces (cut along the x - y plane) for several bulk dopings. The last panel shows, for comparison, the 2D Fermi surface for a 2DEG of density $\sim 10^{14} \text{ cm}^{-2}$ (hybridization between different Fermi surface sheets is not included), deduced from a tight-binding model¹². Colours indicate the character of each Fermi surface sheet along k_y . **d**, Quantum well states, or

their symmetry characters, discussed further below. We note that this band structure agrees with the one depicted in Fig. 1e.

We now quantitatively analyse the observed band structure. The upper and lower parabolic bands have bandwidths of 210 and 100 meV and Fermi momenta of 0.21 and 0.13 \AA^{-1} (best seen in Fig. 3), respectively. Thus, they correspond to light carriers with effective masses $m_y^* \approx 0.7m_e$ along k_y (m_e is the free electron mass). The upper shallow band (Figs 2g, h and 3) has a bandwidth of 40 meV and a Fermi momentum of $k_F \approx 0.3\text{--}0.4 \text{ \AA}^{-1}$, corresponding to heavy carriers with $m_y^* \approx 10m_e\text{--}20m_e$. The lower shallow band (Figs 2i and 3) disperses from about -160 meV at Γ to about -120 meV at the Brillouin zone boundary, and is thus fully occupied. All the above figures differ by less than 10% across all the studied dopings.

As shown in Fig. 4, the resulting Fermi surface consists of a circle of radius 0.21 \AA^{-1} and two ellipsoids, along k_x and k_y , with semi-axes of 0.13 and $0.3\text{--}0.4 \text{ \AA}^{-1}$, respectively. From the area, A_F , enclosed by each Fermi surface, the corresponding 2D carrier density is $n_{2D} = A_F/2\pi^2$. Accounting for the three bands that cross E_F , we find 0.33 ± 0.03 electrons per a^2 (or about $2 \times 10^{14} \text{ cm}^{-2}$), where a is the cubic lattice parameter.

We now rationalize the observed electronic states as the subbands of a 2DEG confined within a few unit cells at the surface of SrTiO₃. To this end, we consider a potential, V_0 , at the surface that confines the

subbands, resulting from the confinement of electrons near the surface of SrTiO₃. The inset shows a wedge-like potential created by an electric field of strength F at the surface, which we use as a simple model to analyse the ARPES data (see main text). The size, L , of the confined 2DEG can be estimated from the extension of the highest occupied state ($n = 2$ in our case) or approximately from the energy difference between the lowest ($n = 1$) and highest occupied states. **e**, Additional degeneracy lifts at Γ occur as a result of spin-orbit coupling, tetragonal and orthogonal distortions, or possible surface reconstructions. This subband hierarchy is the one that best represents the experimental results.

electronic motion along z (Fig. 1d, inset). This lowers the energy of the bands by about V_0 , similar to the ‘band bending’ in semiconductor heterostructures, and produces an energy splitting between the different eigenstates that is inversely proportional to their effective masses along z (m_z^*). The resulting subband structure, depicted in Fig. 1d, consists on a single d_{xy} -like band and two d_{xz}/d_{yz} bands that are degenerate at Γ . As the d_{xy} band has a very large m_z^* , the attractive confining potential will merely pull it below E_F (its energy-split eigenstates will have a negligible separation). Thus, we identify this band with the lower parabolic band in our spectra of Fig. 2, and denote it $E_1(d_{xy})$. By contrast, the d_{xz} and d_{yz} subbands, which are light along z , will show large energy splittings. They are noted $E_n(d_{xz/yz})$ ($n = 1, 2, \dots$) in Fig. 1d.

Other effects beyond this simplified analysis, such as spin-orbit coupling (which has been reported at the LaAlO₃/SrTiO₃ interface²⁰) and/or the low-temperature tetragonal and possibly orthorhombic distortions^{21,22}, can lift the degeneracy between the d_{xz} and d_{yz} subbands, as illustrated in Fig. 1e. This would induce a coupling, resulting in hybridization between the light and heavy subbands, as indeed evidenced by our data for the lower parabolic band and the shallow bands (Fig. 3, linear vertical polarization). Another possibility would be a surface reconstruction, although this is not suggested by our data, which follows the periodicity of the unreconstructed bulk lattice without apparent

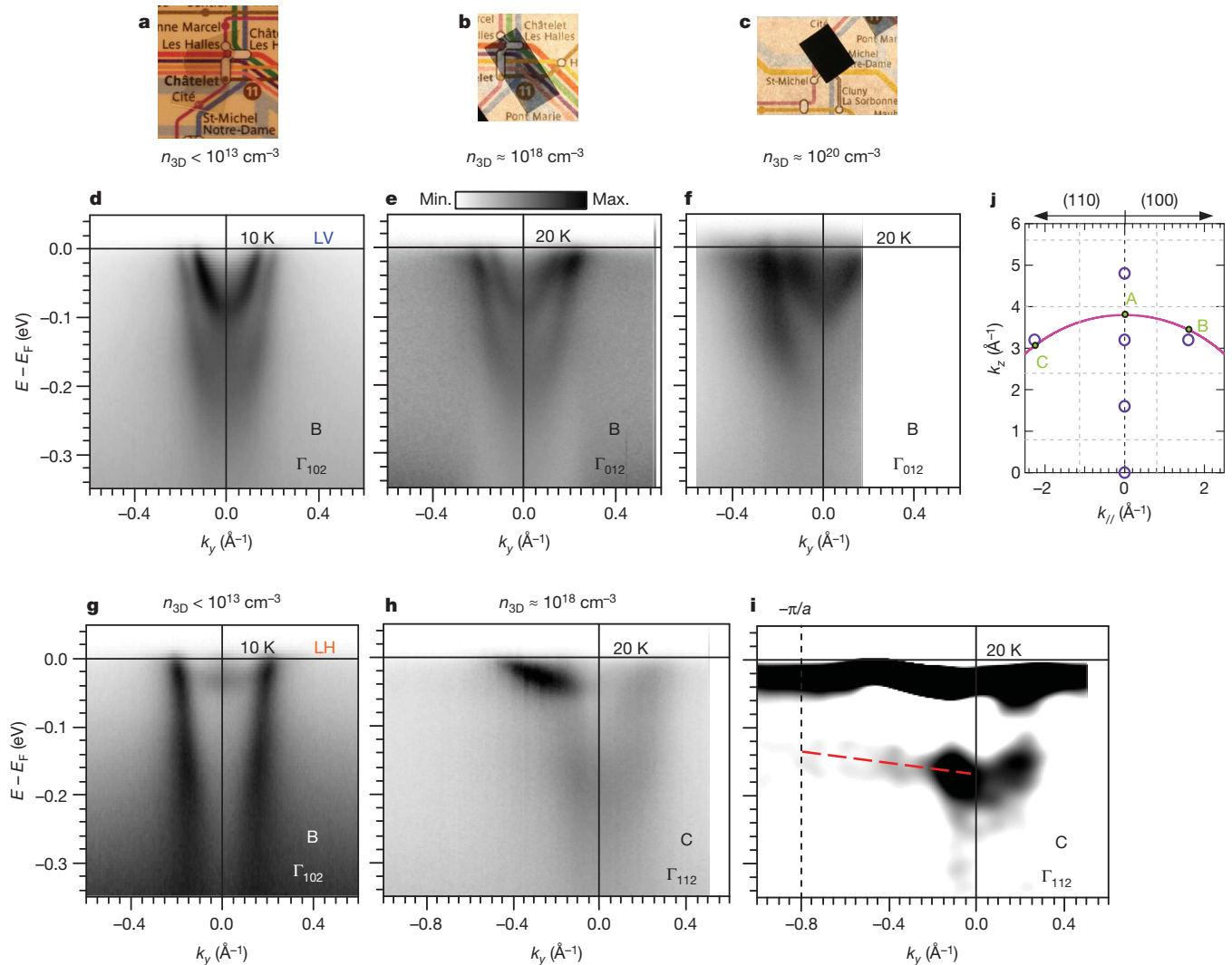


Figure 2 | Universal electronic structure at the surface of SrTiO_3 .

a–c, Photographs and bulk dopings of the samples studied using ARPES. **d–f**, Corresponding energy–momentum intensity maps close to the Γ point (B-like points in **j**). In **d–f**, only the upper and lower parabolic bands are best observed, owing to dipole selection rules for the photon polarizations and measurement geometries used (Methods and Supplementary Information, section IV). If the photon polarization is rotated from linear vertical (LV; **d**) to linear horizontal (LH; **g**), the upper shallow band appears and the upper parabolic band disappears, as illustrated here for the non-doped sample. The effects of collecting electrons around a Γ point at a larger emission angle in a

band folding. Resolving the detailed origin of such a degeneracy lift, which is not relevant to the discoveries reported here, will require future investigations. Hence, we identify the lower shallow band and the upper parabolic band in our spectra (Fig. 2) with the components of the split doublet $E_1(d_{xz})$ and $E_1(d_{yz})$, respectively. From the data, the doublet splitting is $\Delta \approx 60 \text{ meV}$. We finally identify the upper shallow band with $E_2(d_{xz})$. Given that the bottom of this band is at -40 meV , its upper partner, $E_2(d_{yz})$, would occur at about $+20 \text{ meV}$ and cannot be observed.

This subband hierarchy respects the symmetry considerations previously deduced from the effects of light polarization and emission angle, rendering additional support to the picture of a confined 2DEG.

To characterize this 2DEG quantitatively, we use a simple schematic model. We assume that the surface has a homogeneous, positive charge density, generating a confining electric field, F , inside the solid and, thus, a triangular wedge potential of depth V_0 at the surface ($z = 0$) and value $V(z) = -V_0 + eFz$ (e , elementary charge) along z , as depicted in the inset of Fig. 1d. The details of the calculations are

presented in Supplementary Information (section VI). From the experimental splitting, $E_2(d_{xz}) - E_1(d_{xz}) = 120 \text{ meV}$, between the first two subbands of the d_{xz} orbitals, and their light mass, $m_z^* \approx 0.7m_e$, along z , we deduce a strong confining field $F \approx 83 \text{ MV m}^{-1}$. From $E_1(d_{yz})$, the confining potential is estimated to be $V_0 \approx -260 \text{ meV}$. The deduced values of V_0 and F , and the measured effective masses and doublet splitting of 60 meV between the d_{xz} and d_{yz} subbands, suffice for the calculation of all the observed subband energies. The width, L , of the confined 2DEG results from the spatial extension of the highest occupied state, $E_2(d_{xz})$, or, independently, from the approximate expression $eFL \approx E_2(d_{xz}) - E_1(d_{xz})$ (Fig. 1d, inset), yielding consistent values in the range $L \approx 14.5\text{--}18 \text{ \AA}$, or $\sim 4\text{--}5$ unit cells. Additionally, from the confining field, F , and the polarizability of the medium, we calculate an induced surface charge density of ~ 0.25 electrons per \AA^2 , consistent with the experimental value.

The band bending due to confinement should induce an energy downshift of the oxygen $2p$ valence band of about V_0 , creating a gap with respect to E_F that is larger than the optical gap. This effect,

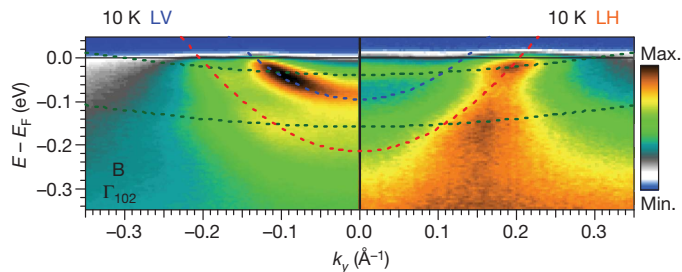


Figure 3 | Summary of subbands for the 2DEG at the surface of SrTiO₃. Side-by-side representation of the LV and LH spectra around Γ_{102} for the non-doped sample. An elongated aspect ratio has been used to discern the shallow bands better. The dotted lines are tight-binding representations of the bands¹², following the same colour scheme as in Fig. 1e. Hybridizations between the lower parabolic band and the shallow bands are observed on the spectra taken with LV polarization.

observed in our data (Supplementary Information, section V), was previously reported and studied in ref. 17, where it was demonstrated that when the cleaved surface of SrTiO₃ is exposed to oxygen, the oxygen 2p valence band shifts by about 200 meV towards E_F and the quasiparticle peak at E_F , which corresponds to our 2DEG, disappears. Surface oxygen vacancies therefore seem to be the most reasonable candidates to explain the origin of the 2DEG observed in our experiments. In fact, cleaving (or fracturing) is likely to produce a massive removal of surface oxygen, much larger than and basically independent of the bulk concentration of vacancies, providing two dopant electrons per created vacancy. These two electrons will delocalize within the potential wedge created by the positively charged layer of surface vacancies. We note that fracturing this almost cubic oxide system necessitates overcoming the strong binding electric forces between atoms, which is equivalent to inducing a ‘mechanical dielectric breakdown’. According to our data analysis, the large electric field generated at the surface of SrTiO₃ is of the order of typical dielectric breakdown fields²³, and is thus compatible with a cleaved surface in this material.

To give additional support to our picture and data analyses, we performed *ab initio* calculations including oxygen vacancies at the surface. The results, presented in the Supplementary Information (section VII), are consistent with the above simplified wedge-potential model for the two types (TiO₂ and SrO) of possible surface termination.

Strikingly, the sheet carrier density, the confinement size, the confining field, F , and potential, V_0 , and the presence of both mobile and quasi-localized carriers in this novel electron gas at the surface of SrTiO₃ compare well with the characteristics of other 2DEGs at different types of SrTiO₃ interface^{6,8,9,12,13,24–27}. Furthermore, our experiments directly show the emergence, at the surface of SrTiO₃, of a metallic subband structure where the titanium $3d_{xy}$ states of light carriers become the first available levels. This situation is similar to that reported in LaAlO₃/SrTiO₃ interfaces²⁸. These remarkable consistencies suggest that all these 2DEGs may be understood on a common basis. In particular, we note that fundamentally different electron confinement mechanisms, such as the electric field created by a polar LaAlO₃ layer⁶, the direct application of an electric field⁹ or the charge leakage between a Mott insulator and a band insulator¹¹, would recreate a 2DEG confined within a nanometric layer of SrTiO₃ with a subband electronic structure similar in nature to that revealed by our measurements.

Our discoveries demonstrate that a highly mobile 2DEG can be readily obtained without any elaborate interface preparation at the surface of SrTiO₃, even in the case of the non-doped 3 eV bandgap insulator. This 2DEG is thus decoupled from the sea of bulk electrons, and constitutes a model system for fundamental and applied research in oxide surfaces/interfaces. An exciting perspective is that other hard-to-cleave complex oxides with multifunctional behaviour, such as multiferroics, ferroelectrics or manganites, might realize similar 2DEGs that inherit the functionalities of the bulk material. Candidates to study in this regard would

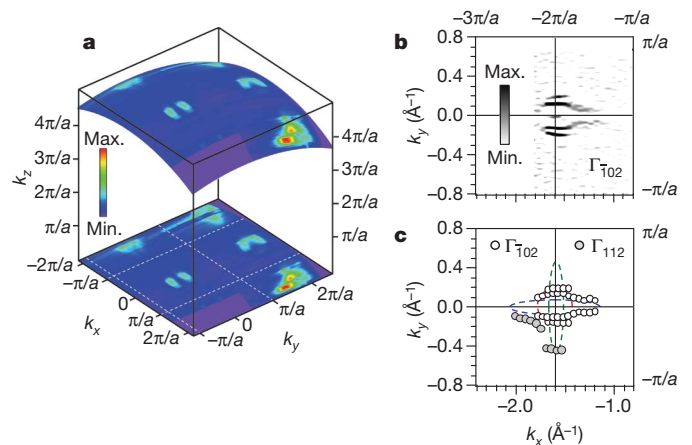


Figure 4 | Fermi surface of 2DEG at the surface of SrTiO₃. **a**, ARPES Fermi surface map (integrated within ± 20 meV around E_F) across several Brillouin zones. The map is displayed on the spherical shell of 3D k space corresponding to the measurement photon energy $h\nu = 47$ eV (Methods), and is also projected on the k_x - k_y plane. Owing to dipole selection rules for the measurement geometry used (Supplementary Information, section IV), Fermi surface sheets of different orbital characters are more intense in different Brillouin zones. Thus, the Fermi surface circle from the lower (d_{xy} -like) light parabolic band is best observed in the Γ_{002} and Γ_{012} (and equivalent) Brillouin zones, whereas the two elliptical, perpendicular Fermi surfaces from the upper parabolic band (d_{yz}) and the upper shallow band (d_{xz}) are best observed in off-normal emission (Γ_{012} , Γ_{112} and equivalent). **b**, Second derivative (negative values only) of the Fermi surface map over Γ_{102} , showing portions of the circular and the elliptical-horizonal Fermi surfaces. The intensity of the vertical ellipse is very weak here owing to dipole selection rules. **c**, Collected Fermi momenta from high-statistics measurements around Γ_{102} (white circles) and Γ_{112} (grey circles), and comparison with the Fermi surfaces resulting from the tight-binding bands of Fig. 3 (colours indicate the character along k_y and follow the scheme used in Fig. 1c–e). All panels correspond to measurements on the low-doped sample.

be other titanates, such as the ferroelectric BaTiO₃ or the Mott-Hubbard insulator LaTiO₃. The properties of such 2DEGs might be tailored by surface doping with either electron-donor or -acceptor species, something that is not always feasible in the bulk, opening a window to explore new phenomena. This simple approach has been applied successfully to tune the density of states in, for example, the surface states of topological insulators²⁹. Our discoveries also raise the appealing possibility that oxygen vacancies may be induced to aggregate or self-organize at the surfaces or interfaces of SrTiO₃ or other oxides—a situation that might be at the origin of the reversible patterning of conducting nano-lines in LaAlO₃/SrTiO₃^{2,7,30}. More generally, we suggest that the nano-engineering of atomic vacancies at the surfaces of transition-metal oxides could provide model systems for low-dimensional, confined electron gases, opening a new avenue in correlated-electron surface science.

METHODS SUMMARY

High-quality, single crystals of SrTiO₃ doped with oxygen vacancies (except for the non-doped sample) were measured throughout this work. The ARPES experiments were done at the Synchrotron Radiation Center (University of Wisconsin-Madison) and at the Synchrotron SOLEIL (France), using different photon energies, intensities, polarizations and measurement geometries. In all cases, the angular and energy resolutions were 0.25° and 20 meV, respectively. The samples were fractured *in situ* along the c axis below 20 K at a pressure lower than 6×10^{-11} torr. Using calibrated photodiodes, we checked that the beam intensity per unit area at the same photon energy differs by at least one decade between the two synchrotrons. As the observed 2DEG is identical in both cases, this indicates that its subbands are not significantly populated by photo-carrier injection. The results have been reproduced in at least two samples for each bulk doping. The methods associated with this work are detailed in Methods and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 July; accepted 26 November 2010.

- Ramirez, A. P. Oxide electronics emerge. *Science* **315**, 1377–1378 (2007).
- Cen, C., Thiel, S., Mannhart, J. & Levy, J. Oxide nanoelectronics on demand. *Science* **323**, 1026–1030 (2009).
- Ohtomo, A. & Hwang, H. Y. A high-mobility electron gas at the $\text{LaAlO}_3/\text{SrTiO}_3$ heterointerface. *Nature* **427**, 423–426 (2004).
- Ahn, C. H., Triscone, J.-M. & Mannhart, J. Electric field effect in correlated oxide systems. *Nature* **424**, 1015–1018 (2003).
- Ohtomo, A., Muller, D. A., Grazul, J. L. & Hwang, H. Y. Artificial charge-modulation in atomic-scale perovskite titanate superlattices. *Nature* **419**, 378–380 (2002).
- Thiel, S., Hammer, G., Schmehl, A., Schneider, C. W. & Mannhart, J. Tunable quasi-two-dimensional electron gases in oxide heterostructures. *Science* **313**, 1942–1945 (2006).
- Cen, C. *et al.* Nanoscale control of an interfacial metal–insulator transition at room temperature. *Nature Mater.* **7**, 298–302 (2008).
- Reyren, N. *et al.* Superconducting interfaces between insulating oxides. *Science* **317**, 1196–1199 (2006).
- Ueno, K. *et al.* Electric-field-induced superconductivity in an insulator. *Nature Mater.* **7**, 855–858 (2008).
- Brinkman, A. *et al.* Magnetic effects at the interface between non-magnetic oxides. *Nature Mater.* **6**, 493–496 (2007).
- Okamoto, S. & Millis, A. J. Electronic reconstruction at an interface between a Mott insulator and a band insulator. *Nature* **428**, 630–633 (2004).
- Popović, Z. S., Satpathy, S. & Martin, R. M. Origin of the two-dimensional electron gas carrier density at the LaAlO_3 on SrTiO_3 interface. *Phys. Rev. Lett.* **101**, 256801 (2008).
- Basletic, M. *et al.* Mapping the spatial distribution of charge carriers in $\text{LaAlO}_3/\text{SrTiO}_3$ heterostructures. *Nature Mater.* **7**, 621–625 (2008).
- Nakamura, H. *et al.* Tuning of metal–insulator transition of two-dimensional electrons at parylene/ SrTiO_3 interface by electric field. *J. Phys. Soc. Jpn* **78**, 083713 (2009).
- Mattheiss, L. F. Energy bands for KNiF_3 , SrTiO_3 , KMoO_3 and KTaO_3 . *Phys. Rev. B* **6**, 4718–4740 (1972).
- Fujimori, A. *et al.* Doping-induced changes in the electronic structure of $\text{La}_{1-x}\text{Sr}_x\text{TiO}_3$: limitation of the one-electron rigid-band model and the Hubbard model. *Phys. Rev. B* **46**, 9841–9844 (1992).
- Aiura, Y. *et al.* Photoemission study of the metallic state of lightly electron-doped SrTiO_3 . *Surf. Sci.* **515**, 61–74 (2002).
- Ishida, Y. *et al.* Coherent and incoherent excitations of electron-doped SrTiO_3 . *Phys. Rev. Lett.* **100**, 056401 (2008).
- Chang, Y. J., Bostwick, A., Kim, Y. S., Horn, K. & Rotenberg, E. Structure and correlation effects in semiconducting SrTiO_3 . *Phys. Rev. B* **81**, 235109 (2010).
- Caviglia, A. D. *et al.* Tunable Rashba spin-orbit interaction at oxide interfaces. *Phys. Rev. Lett.* **104**, 126803 (2010).
- Lytle, F. W. X-ray diffractometry of low-temperature phase transformations in strontium titanate. *J. Appl. Phys.* **35**, 2212–2215 (1964).
- Mattheiss, L. F. Effect of the 110°K transition on the SrTiO_3 conduction bands. *Phys. Rev. B* **6**, 4740–4753 (1972).
- Barret, H. H. Dielectric breakdown of single-crystal strontium titanate. *J. Appl. Phys.* **35**, 1420–1425 (1964).
- Copie, O. *et al.* Towards two-dimensional metallic behavior at $\text{LaAlO}_3/\text{SrTiO}_3$ interfaces. *Phys. Rev. Lett.* **102**, 216804 (2009).
- Seo, S. S. A. *et al.* Multiple conducting carriers generated in $\text{LaAlO}_3/\text{SrTiO}_3$ heterostructures. *Appl. Phys. Lett.* **95**, 082107–082109 (2009).
- Sing, M. *et al.* Profiling the interface electron gas of $\text{LaAlO}_3/\text{SrTiO}_3$ heterostructures with hard X-ray photoelectron spectroscopy. *Phys. Rev. Lett.* **102**, 176805 (2009).
- Pentcheva, R. & Pickett, W. E. Ionic relaxation contribution to the electronic reconstruction at the n-type $\text{LaAlO}_3/\text{SrTiO}_3$ interface. *Phys. Rev. B* **78**, 205106 (2008).
- Salluzzo, M. *et al.* Orbital reconstruction and the two-dimensional electron gas at the $\text{LaAlO}_3/\text{SrTiO}_3$ interface. *Phys. Rev. Lett.* **102**, 166804 (2009).
- Hsieh, D. *et al.* A tunable topological insulator in the spin helical Dirac transport regime. *Nature* **460**, 1101–1105 (2009).
- Xie, Y., Bell, C., Yajima, T., Hikita, Y. & Hwang, H. Y. Charge writing at the $\text{LaAlO}_3/\text{SrTiO}_3$ surface. *Nano Lett.* **10**, 25882591 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to N. Bontemps, R. Claessen, Y. Fagot-Revurat, M. Gabay, I. C. Infante, D. Malterre, A. J. Millis and F. Reinert for discussions, to E. Jacquet for help with the sample preparation and to R. Guerrero for help with the transport measurements. This work was supported by the ANR OXITRONICS and the CNRS-CSIC PICS ‘POSTIT’ project under grant number PICS2008FR1. The Synchrotron Radiation Center, University of Wisconsin-Madison, is supported by the National Science Foundation under award no. DMR-0537588. The Ames Laboratory is operated for the US DOE by Iowa State University under contract number W-7405-ENG-82. R.W. is a research fellow of CONICET-Argentina, supported by CONICET (grant PIP 112-200801-00047) and ANPCyT grant PICT 837/07. X.G.Q. is supported by the MOST and NSF of China, and G.H. is supported by the Spanish Government under project numbers MAT2008-06761-C03 and NANOSELECT CSD2007-00041.

Author Contributions A.F.S.-S. and O.C. contributed equally to this work, from project conception and ARPES measurements to data analysis, interpretation and writing of the manuscript. The contributions of other authors are as follows. Project conception: G.H., M.B., A.B., M.J.R.; ARPES measurements: T.K., F.F., S.P., F.B., A.N.; infrastructure for ARPES experiments at SOLEIL: F.B., A.T.-I., P.L.F.; samples: X.G.Q., G.H., M.B., Y.A., P.L., A.B.; transport measurements: N.R., Y.A.; data analysis, interpretation, slab-LDA calculations: R.W., M.J.R.; input to writing the manuscript: M.J.R. All authors extensively discussed the results and the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.F.S.-S. (andres.santander@csnsm.in2p3.fr).

METHODS

Sample preparation and measurement technique. High-quality, single crystals of SrTiO_3 were doped with oxygen vacancies by high temperature treatment in low oxygen pressure, except for the non-doped sample, which was a bare SrTiO_3 substrate measured without any previous preparation. The bulk carrier concentrations were deduced from Hall and resistivity measurements (Supplementary Information, section II).

The ARPES experiments were done at the Synchrotron Radiation Center (SRC, University of Wisconsin-Madison) and at the Synchrotron SOLEIL (France), using 47 eV (SRC) and 45 eV (SOLEIL) linearly polarized photons, a Scienta 2002 detector with horizontal slits for the highly doped sample (SRC), and a Scienta R4000 with vertical slits for the low-doped (SRC) and non-doped (SOLEIL) samples. The above photon energies, close to resonant photoemission at the titanium $3p-3d$ edge, yield intense quasiparticle peaks in SrTiO_3 (ref. 17). The momentum and energy resolutions were 0.25° and 20 meV, respectively. The mean diameter of the incident photon beam was less than $50\text{ }\mu\text{m}$ at the Synchrotron SOLEIL and about $300\text{ }\mu\text{m}$ at the SRC. We checked, using calibrated photodiodes, that the photon beam intensity per unit area at the above photon energies differs by at least one order of magnitude between SOLEIL and the SRC. As the observed subbands are the same in both cases, this indicates that the 2DEG is the intrinsic stable state of the ultrahigh-vacuum fractured surface of SrTiO_3 , and that its subbands are not significantly populated by photo-carrier injection.

The samples were fractured *in situ* along the c axis at 20 K (SRC) or 10 K (SOLEIL), at a pressure lower than 6×10^{-11} torr. After fracturing, one narrow, highly emitting region was selected and kept fixed for each set of measurements. Supplementary Information (section I) presents details of the fracturing procedure and a discussion of the surface terminations in relation to the data shown here.

The Fermi momenta and bandwidths were determined from the peak positions of the momentum distribution curves at $E = 0$ and the energy distribution curves at $k = 0$, respectively.

The results have been reproduced in at least two samples for each bulk doping. **Measurement geometries.** The sample's surface defines the x - y plane. The crystals were oriented using Laue diffraction and mounted with the (010) axis along the angle-resolved direction of the detector, defining the k_y direction. The photon energy fixes the radius of the k -space hemisphere for the measurements, which was $\sim 3.8\text{ }\text{\AA}^{-1}$ (or about $0.7\pi/a$ above Γ_{002}) for SrTiO_3 at photon energies around 47 eV (refs 17, 18). The data in Figs 2d, g and 3 were collected from B-like points slightly above Γ_{102} , with the photon polarization along y (LV) in Fig. 2d and in the x - z plane (LH) in Fig. 2h. The data in Figs 2e, f correspond to B-like points slightly above Γ_{012} and those in Figs 2h, i correspond to C-like points slightly below Γ_{112} . For these points, the photon polarization is not parallel to any of the sample's symmetry directions or planes, and the spectra have different symmetry mixtures. The data in Fig. 4 were collected with LH photon polarization. Supplementary Information (section IV) discusses further the measurement geometries and their corresponding dipole selection rules.

Nanoscale chemical tomography of buried organic–inorganic interfaces in the chiton tooth

Lyle M. Gordon¹ & Derk Joester¹

Biological organisms possess an unparalleled ability to control the structure and properties of mineralized tissues. They are able, for example, to guide the formation of smoothly curving single crystals or tough, lightweight, self-repairing skeletal elements¹. In many biominerals, an organic matrix interacts with the mineral as it forms, controls its morphology and polymorph, and is occluded during mineralization^{2–4}. The remarkable functional properties of the resulting composites—such as outstanding fracture toughness and wear resistance—can be attributed to buried organic–inorganic interfaces at multiple hierarchical levels⁵. Analysing and controlling such interfaces at the nanometre length scale is critical also in emerging organic electronic and photovoltaic hybrid materials⁶. However, elucidating the structural and chemical complexity of buried organic–inorganic interfaces presents a challenge to state-of-the-art imaging techniques. Here we show that pulsed-laser atom-probe tomography reveals three-dimensional chemical maps of organic fibres with a diameter of 5–10 nm in the surrounding nano-crystalline magnetite (Fe_3O_4) mineral in the tooth of a marine mollusc, the chiton *Chaetopleura apiculata*. Remarkably, most fibres co-localize with either sodium or magnesium. Furthermore, clustering of these cations in the fibre indicates a structural level of hierarchy previously undetected. Our results demonstrate that in the chiton tooth, individual organic fibres have different chemical compositions, and therefore probably different functional roles in controlling fibre formation and matrix–mineral interactions. Atom-probe tomography is able to detect this chemical/structural heterogeneity by virtue of its high three-dimensional spatial resolution and sensitivity across the periodic table. We anticipate that the quantitative analysis and visualization of nanometre-scale interfaces by laser-pulsed atom-probe tomography will contribute greatly to our understanding not only of biominerals (such as bone, dentine and enamel), but also of synthetic organic–inorganic composites.

Organic scaffolds in biominerals generally comprise a fibrous structural element, for example, collagen in bone⁷ or chitin in molluscs and crustaceans⁸. Acidic macromolecules, such as the non-collagenous proteins in bone⁹ and Asprich proteins in the bivalve prismatic layer¹⁰, are usually associated with the fibrous scaffold. Although these proteins may be intrinsically disordered, their charge predisposes them to interact with the forming mineral¹¹. They are thought to have a key role in controlling crystal polymorph and orientation. Additionally, ions such as Mg^{2+} (refs 12 and 13) and polyphosphates¹⁴ are known to be important modulators of mineralization processes. Consequently, our understanding of biological control over mineral growth and the resulting functional properties depends critically on characterizing buried organic–inorganic interfaces.

Characterization of the organic scaffold frequently requires wet chemical demineralization and drying that can result in structural artefacts and loss of diffusive species. The removal of either mineral or organic material also renders it impossible to analyse the interface between the two. The structural complexity of buried interfaces in a synthetic calcite single crystal with occluded agarose fibres was only

recently demonstrated by scanning transmission electron microscopy (STEM) and tomography¹⁵. However, even this advanced imaging modality does not reveal the chemical complexity of the sample in terms of the localization of different chemical species in the scaffold or the mineral. Although high-resolution elemental mapping by electron microscopy is in principle possible, resolution and sensitivity are greatly limited by the susceptibility of biological materials to beam damage.

We demonstrate herein that atom-probe tomography (APT), an established technique in metallurgical and semiconductor research^{16,17}, is capable of chemical tomography of nanoscale buried organic–inorganic interfaces. In particular, we investigate the nano-crystalline magnetite cap (Fe_3O_4 , cubic space group $Fd\bar{3}m$; Fig. 1) of the chiton tooth, which is a classical model system for the study of matrix-mediated mineralization¹. Chitons (also known as sea cradles) are marine molluscs of the class Polyplacophora. Chiton teeth are arranged in rows along the radula (rasping tongue). Similar to vertebrate teeth, mature chiton teeth consist of a softer core (apatite or iron phosphate) capped by a hard magnetite layer¹. Depending on the species, the capping may cover the whole tooth or just the leading edge, and other mineral phases (for example, lepidocrocite) may be present. The outstanding fracture toughness and wear resistance of the tooth results from the organic–inorganic interfaces over multiple levels of hierarchy, which deflect and arrest cracks¹⁸. Remarkably, the radula acts as a conveyor-belt on which an organic matrix scaffold composed of semi-crystalline α -chitin (poly- β -1,4-*N*-acetylglucosamine) and protein is first deposited¹⁹. As the maturing tooth progresses along the radula (about one row per day), the scaffold is remodelled and filled in with mineral, which occludes the organic matrix in the process. All stages of tooth development are thus present in one animal.

Samples for transmission electron microscopy (TEM) and APT were prepared from polished sagittal cross-sections of teeth of the Eastern Beaded Chiton (*Chaetopleura apiculata*, Fig. 2). Focused ion beam (FIB) lift-out techniques were used to prepare electron-transparent

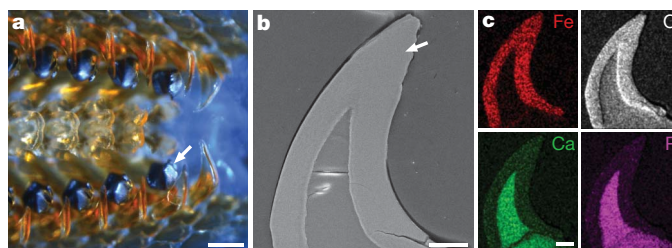


Figure 1 | Chiton radula and tooth structure. **a**, Reflected light optical micrograph of the tip of a *C. apiculata* radula, with four rows of fully mineralized teeth (arrow). Scale bar, 200 μm . **b**, SEM image of polished cross-section of a tooth, with approximate site where samples for APT and TEM were extracted (arrow). Scale bar, 50 μm . **c**, Scanning electron microscopy-energy dispersive X-ray spectroscopy (SEM-EDS) elemental maps of cross-section (b) showing a Ca/P/O-rich core (apatite) and Fe/O-rich cap (magnetite). Scale bar, 20 μm .

¹Northwestern University, Department of Materials Science and Engineering, 2220 Campus Drive, Evanston, Illinois 60208, USA.

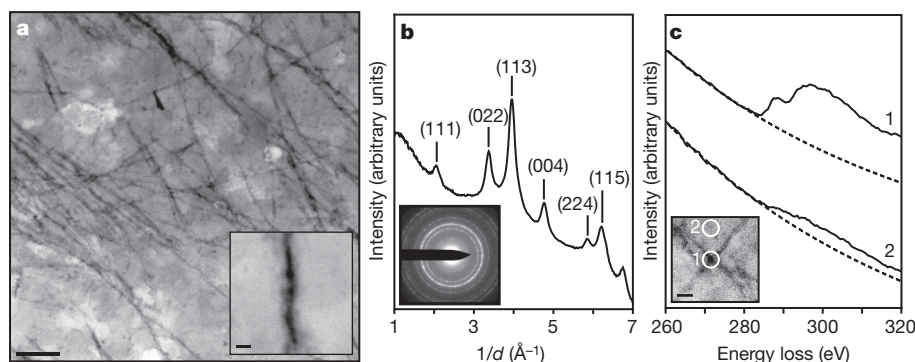


Figure 2 | Chiton tooth magnetite and occluded organic fibres. **a**, High-angle annular dark-field STEM image of chiton tooth magnetite showing dark (low- Z) fibrous structures with a length $>1\ \mu\text{m}$ and a diameter of $\sim 5\text{--}10\ \text{nm}$ embedded in the nanocrystalline magnetite (for low angle annular dark field STEM see Supplementary Fig. 7). Scale bar, 100 nm. The inset shows detail of the fibre. Scale bar, 10 nm. **b**, Radially integrated selected-area electron diffraction pattern from the chiton tooth cusp; labelled planes correspond to

magnetite. The inset shows the two-dimensional selected-area electron diffraction pattern. **c**, Carbon-K edge electron energy-loss spectra recorded at fibre intersection (1) and off-fibre (2) demonstrate that fibres are carbon-rich and probably correspond to the chitin observed in unmineralized and demineralized chiton teeth. The inset shows the locations where the measurements were acquired. Scale bar, 20 nm.

samples from the leading edge of the tooth cap. The identity of the mineral (magnetite) was confirmed by selected-area electron diffraction. High-angle annular dark-field STEM reveals numerous dark (weakly scattering) fibrous structures on a bright (strongly scattering) background with dimensions ($d = 7.6 \pm 2.4\ \text{nm}$) similar to those observed in de-mineralized chiton teeth (Fig. 2)¹⁹. The carbonaceous nature of the fibres was confirmed by carbon K-edge electron energy-loss spectroscopy.

An atom probe is a point projection microscope where, in the presence of a very high electric field, individual or molecular ions sequentially field-evaporate from a sharp tip and are projected onto a two-dimensional position-sensitive detector¹⁶. The mass-to-charge ratio (m/z) and thus the chemical identity of each ion are determined by time-of-flight mass spectrometry using timed picosecond laser pulses to trigger evaporation events. The sequence and location of ions impinging on the detector enables the reconstruction of the three-dimensional structure of the sample. APT analyses volumes of the order of $10^5\ \text{nm}^3$ with subnanometre spatial resolution¹⁶. The recent development of ultraviolet laser pulsing greatly increases the scope of the technique, to include high-resistivity materials and organics.

In APT mass spectra of samples from the chiton tooth cusp, magnetite-derived iron, oxygen and iron oxide molecular ions prevail (Table 1), similar to spectra of geological magnetite (Supplementary Fig. 2 and ref. 20). As is typical for oxides, about 25% of the oxygen evaporates as O_2^{2+} , which cannot be differentiated from O^+ (refs 20–22). The measured Fe/O ratio for both the biological (0.97 ± 0.15) and geological magnetite (1.13 ± 0.08) is thus higher than the theoretical value of 0.75. APT mass spectra also contain molecular fragments with m/z values corresponding to C/N/O-containing ions (Table 1 and Fig. 3), which are not present in spectra of geological magnetite. Although ions such as C^+ and C^{2+} are also found in metal carbides, there is no precedence for carbide formation under physiological conditions. In addition, other ions frequently associated with carbides, such as C_n^+ , C_n^{2+} ($n > 1$) and FeC^+ were not detected. However, samples prepared from chitin thin films yield a similar series of molecular ions, confirming that these ions originate from the organic scaffold within the tooth (Fig. 3 and Supplementary Fig. 3). While there is some overlap between organic and mineral-derived ions, for example, $^{12}\text{C}^{16}\text{O}_2^+$ and $^{56}\text{Fe}^{16}\text{O}_2^{2+}$ at $m/z \approx 44$,

analysis based on natural isotopic abundances allows identification and quantification (see Supplementary Notes).

Demonstrating the unique sensitivity of APT across the entire periodic table, APT spectra reveal the presence of the light elements sodium and magnesium and trace amounts of manganese. Although the role of Na in biomineralization is not well understood, Mg^{2+} has been shown to modulate mineralization processes, and Mg-binding motifs have been identified in acidic proteins^{13,23}. Mn, on the other hand, is a highly regulated, essential cofactor in many redox-active enzymes²⁴ and may play a part in magnetite mineralization.

We used standard reconstruction algorithms to generate three-dimensional tomograms (Fig. 4 and Supplementary Movies 1–3). The organic-matter-derived ions are clearly seen to originate from fibres occluded within the mineral. The total amount of carbon is less than what would be expected from crystalline chitin fibres. This may be a consequence of overlapping peaks in time-of-flight spectra, preferential evaporation, or biological remodelling of fibres during mineralization²⁵. Nevertheless, the fibre diameter (5–10 nm) closely resembles the organic fibres observed in STEM images of the mineralized tooth (Fig. 2a). Furthermore, Na^+ and Mg^{2+} clearly co-localize with organic fibres (Fig. 4). Surprisingly, no biologically relevant counter ions (for example, chloride, phosphate, sulphate) could be detected. Given that chitin is neutral (or, if partially deacetylated, slightly positively charged), the most likely scenario is that acidic proteins associated with the chitin scaffold bind the cations. Interestingly, within each fibre there appear to be small domains around which cations cluster (Fig. 4). It is conceivable that these domains correspond to crystalline α -chitin domains that exclude acidic proteins and the cations associated with these.

Unexpectedly, in some samples fibres only bind Na^+ (Fig. 4a and Supplementary Movie 1), whereas Mg^{2+} is bound exclusively in others (Fig. 4d and Supplementary Movie 3). This is despite the fact that these samples were taken within a few micrometres of each other. Both Mg^{2+} and Na^+ co-localize with fibres in fibre bundles (Supplementary Movie 2 and Supplementary Fig. 6). There are many examples of proteins that selectively bind and/or transport cations; they frequently differentiate between ions based on their ionic radius, preferred coordination geometry, and hard/soft acid/base properties²⁶. For example, the so-called DEAD box motif found in some acidic proteins from biominerals has been suggested to have a role in Mg^{2+} binding¹². It is thus conceivable that in the chiton tooth scaffold, chitin-binding proteins with specificity for either Na^+ or Mg^{2+} are assembled with chitin into a fibre that shows selective binding. The cations themselves may be integrated as part of the assembly, or come in at a later stage. Although we cannot yet provide accurate statistics of the relative frequency of Mg- and Na-binding fibres, it is clear that this unprecedented chemical heterogeneity requires the extension of the number of hierarchical levels to

Table 1 | Atomic and molecular ions identified in chiton teeth from APT mass spectra

| | Identified ions |
|----------------|---|
| Magnetite | O^+ , O_2^+ , Fe^{2+} , Fe^+ , FeO^{2+} , FeO_2^{2+} , $\text{Fe}_2\text{O}_3^{2+}$, $\text{Fe}_3\text{O}_4^{2+}$, FeO_n^+ , $n = 1\text{--}4$ |
| Organic matrix | C^{2+} , C^+ , CO^+ , CO_2^+ , N^+ , NH^+ , NO^+ , NO_2^+ , CNO^+ |
| Other ions | Mg^{2+} , Na^+ , Mn^{2+} , MnO_n^{z+} , $n = 1\text{--}3$, $z = 1, 2$ |

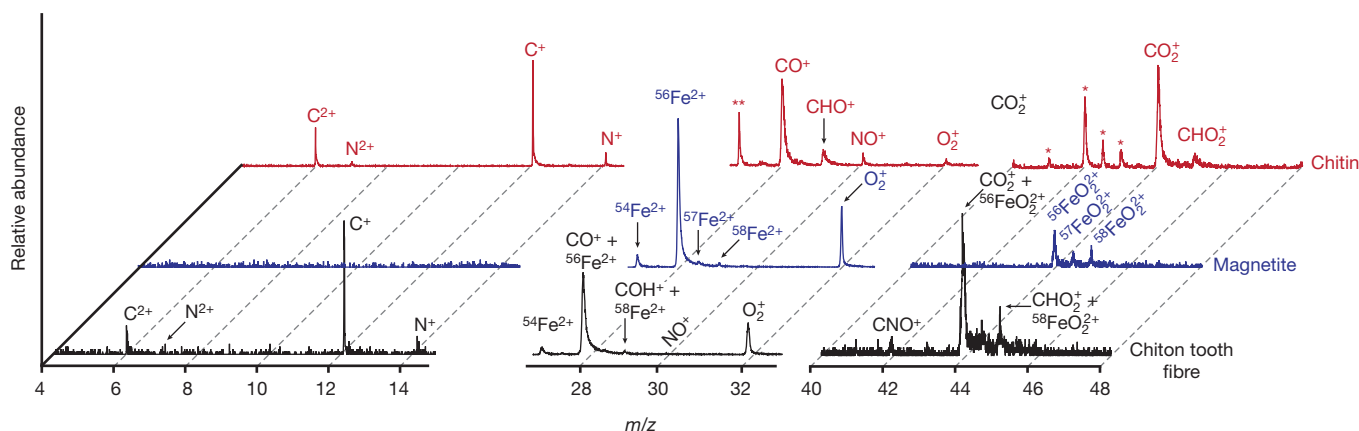


Figure 3 | APT time-of-flight m/z spectra. Typical spectra of organic fibres (black), the surrounding magnetite (blue), and a chitin thin film (red). We note the presence of characteristic chitin-derived atomic and molecular ions (C^+ , C^{2+} , N^+ , N^{2+} , CO^+ , CO_2^+) in the chiton tooth fibre but not in the surrounding magnetite. Spectral overlap between species (for example, CO_2^+ and $^{56}FeO_2^{2+}$)

derived from organics and magnetite can be detected by analysis of natural abundance of Fe isotopes. Species originating from the substrate of the chitin thin film, CrO_2^{2+} (*) and $^{54}Cr^{2+}$ (**), are indicated with asterisks. To account for large differences in relative abundance, each m/z range is scaled separately.

describe the architecture of the chiton tooth accurately. It is possible that similar differences in elemental distributions are present in other biominerals, such as bone or enamel, but have so far been overlooked.

Proximity histograms (proxigrams)²⁷ that give concentration as a function of distance to the fibre–mineral interface (defined by an iso-surface at 1 C nm^{-3} , Supplementary Fig. 4) reveal that the interface is graded rather than sharp, creating a 2–4-nm-wide region of interpenetrating mineral and organic material (Fig. 4). A graded or rough interface may improve mineral/organic adhesion and enhance mechanical load transfer and toughness. Furthermore, this results in an interface fundamentally different from that observed in single crystals of calcite grown in agarose gels, where high-index planes at the interface create a sharp boundary¹⁵. Proxigrams also reveal a roughly twofold increase of the manganese concentration in the fibres (Supplementary Fig. 5). This raises the intriguing possibility that Mn-dependent enzymes, some of which may have chitin-binding domains, catalyse the reductive transformation of ferrihydrite to magnetite.

With the fibre backbone being semi-crystalline α -chitin, it follows that there are crystalline domains and amorphous regions that connect

such domains. It seems reasonable to assume that single strands or loops/coils of amorphous chitin create the rough fibre surface we observe (Fig. 5). The distribution of Na^+ and Mg^{2+} suggests that chitin-binding proteins with selectivity for either or both cations decorate the fibre surface. They may also be present on the inside of the fibre, but are probably excluded from crystalline domains. Fibre–mineral interactions may be mediated by these cations or direct contacts between side-chain carboxylates, amines, and hydroxyl groups and iron cations or oxy-anions on the mineral surface.

On the basis of this model, it is conceivable that fibres decorated with cations with strongly different charge density (that of Mg^{2+} is 4–10 times greater than Na^+ , depending on the coordination number) differ in their functional roles. Modulation of colloidal properties such as the ζ -potential may be an important way to control fibre self-assembly and maintain the fibre diameter such that mineral overgrowth is not impeded. By inducing counter-ion condensation, the fibre chemistry could also influence where and when nucleation of the mineral phase occurs, similar to what has been proposed for nacre organic matrix²⁸. Different strength and geometry of coordinative bonds mediated by

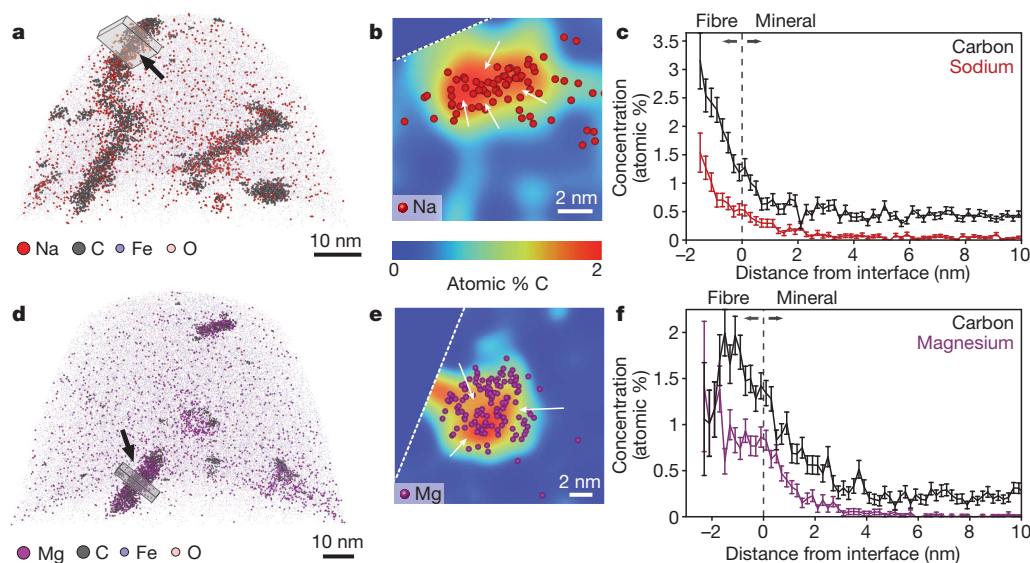


Figure 4 | Three-dimensional reconstructions and proxigrams. **a, d**, Two representative samples containing organic fibres that exclusively bind Na^+ (**a–c**) or Mg^{2+} (**d–f**). For clarity, only about 5% of the Fe/O ions are rendered; the edge of the field-of-view is marked (dashed line in **b**). **b, e**, Overlay of Na^+ (**b**, red spheres) and Mg^{2+} (**e**, magenta) ion positions on carbon concentration

maps integrated over the boxed regions indicated in **a** and **d**. Some regions of the fibres appear devoid of Na or Mg (arrows). **c, f**, Proximity histograms (error bars, $\pm 1\sigma$) of Na/C (**c**) and Mg/C (**f**) across the organic–inorganic interface of fibres indicated by arrows in **a** and **d**. Interfaces appear graded over 2–4 nm.

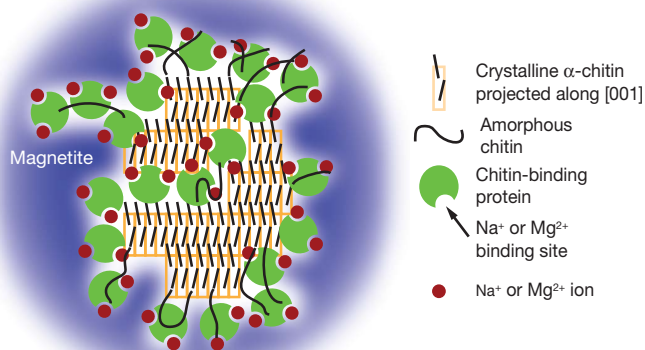


Figure 5 | Model of a chiton tooth organic fibre. The backbone of the fibre consists of semi-crystalline α -chitin. In the crystalline domains, anti-parallel polysaccharide chains are arranged along the [001] direction and are aligned with the fibre axis. Amorphous chitin connects these domains and creates a rough fibre surface. Hypothetical chitin-binding proteins in addition have acidic domains that specifically bind Na^+ or Mg^{2+} . Fibre–mineral interactions may be mediated by these cations and/or acidic, basic or hydroxyl side chains. It is not yet clear whether the magnetite–organic interface involves specific crystallographic planes, is extensively reconstructed, or even strongly disordered (amorphous).

Na^+ and Mg^{2+} will affect interactions within the organic matrix, between mineral and organic matrix and/or control the amount of water retained in the final structure. This could modulate the ability of organic matrix molecules to accommodate large strains by elastic uncoiling ('hidden length'), or control energy dissipation during plastic deformation (that is, toughness) through the strength and number of sacrificial bonds²⁹. Water is an important lubricant, and the incorporation of Mg^{2+} with its tightly bound water molecules may prevent the composite from becoming overly brittle or facilitate reforming sacrificial bonds after deformation, thus contributing to self-healing. Whether the chiton uses these effects to control mineral deposition and properties of the final composite and how we might adapt them to bio-inspired materials synthesis is the subject of our ongoing investigation.

METHODS SUMMARY

Samples for APT and TEM were prepared using *in situ* FIB lift-out procedures. The chiton tooth samples were prepared from a radula extracted from an Eastern Beaded Chiton (*Chaetopleura apiculata*). Sagittal cross-sections of the teeth for FIB were prepared by chemical fixation, resin embedding, mechanical grinding and polishing and sputter coating with Pt. FIB lift-outs were performed on the leading edge of the magnetite tooth cusp. Geological specimen sections were cut with a diamond saw and mechanically polished. Chitin standards were prepared by dissolving purified chitin in hexafluoroisopropanol and spin-coating on a Cr-coated silicon wafer before coating with Ag. Atom probe tomography was conducted using a LEAP 4000XSi (Cameca). Laser pulsing (wavelength $\lambda = 355$ nm, 160–500 kHz, 40–75 pJ per pulse) was used to initiate field evaporation. Direct-current potential on the microtip during APT was controlled to maintain a constant evaporation rate (0.0025 or 0.005 ions per pulse). The base temperature of the sample was maintained at 40 or 60 K. The ambient vacuum pressure was maintained below 10^{-8} Pa. The three-dimensional reconstruction of APT data was performed using published algorithms assuming a hemispherical tip shape and an electric-field dependent tip radius³⁰.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 July; accepted 18 November 2010.

- Lowenstam, H. A. & Weiner, S. *On Biomineralization* (Oxford University Press, USA, 1989).
- Meldrum, F. C. & Colfen, H. Controlling mineral morphologies and structures in biological and synthetic systems. *Chem. Rev.* **108**, 4332–4432 (2008).

- Weiner, S. Biomineralization: a structural perspective. *J. Struct. Biol.* **163**, 229–234 (2008).
- Aizenberg, J. in *Nanomechanics of Materials and Structures* (eds Chuang, T. J., Anderson, P. M., Wu, M. K. & Hsieh, S.) 99–108 (Springer, 2006).
- Fratzl, P. & Weinkamer, R. Nature's hierarchical materials. *Prog. Mater. Sci.* **52**, 1263–1334 (2007).
- Koch, N. Organic electronic devices and their functional interfaces. *ChemPhysChem* **8**, 1438–1455 (2007).
- Weiner, S. & Wagner, H. D. The material bone: structure-mechanical function relations. *Annu. Rev. Mater. Res.* **28**, 271–298 (1998).
- Falini, G. & Fermani, S. Chitin mineralization. *Tissue Eng.* **10**, 1–6 (2004).
- George, A. & Veis, A. Phosphorylated proteins and control over apatite nucleation, crystal growth, and inhibition. *Chem. Rev.* **108**, 4670–4693 (2008).
- Gotliv, B. *et al.* Asprich: a novel aspartic acid-rich protein family from the prismatic shell matrix of the bivalve *Atrina rigida*. *ChemBioChem* **6**, 304–314 (2005).
- Evans, J. S. "Tuning in" to mollusk shell nacre- and prismatic-associated protein terminal sequences. Implications for biomineralization and the construction of high performance inorganic-organic composites. *Chem. Rev.* **108**, 4455–4462 (2008).
- Wang, D., Wallace, A. F., De Yoreo, J. J. & Dove, P. M. Carboxylated molecules regulate magnesium content of amorphous calcium carbonates during calcification. *Proc. Natl Acad. Sci. USA* **106**, 21511–21516 (2009).
- Tao, J. H., Zhou, D. M., Zhang, Z. S., Xu, X. R. & Tang, R. K. Magnesium-aspartate-based crystallization switch inspired from shell molt of crustacean. *Proc. Natl Acad. Sci. USA* **106**, 22096–22101 (2009).
- Omelson, S. J. & Grynpas, M. D. Relationships between polyphosphate chemistry, biochemistry and apatite biomineralization. *Chem. Rev.* **108**, 4694–4715 (2008).
- Li, H., Xin, H. L., Muller, D. A. & Estroff, L. A. Visualizing the 3D internal structure of calcite single crystals grown in agarose hydrogels. *Science* **326**, 1244–1247 (2009).
- Kelly, T. F. & Miller, M. K. Atom probe tomography. *Rev. Sci. Instrum.* **78**, 031101 (2007).
- Seidman, D. N. & Stiller, K. A renaissance in atom-probe tomography. *MRS Bull.* **10**, 717–749 (2009).
- van der Wal, P., Giesen, H. J. & Videler, J. J. Radular teeth as models for the improvement of industrial cutting devices. *Mater. Sci. Eng. C* **7**, 129–142 (2000).
- Evans, L., Macey, D. & Webb, J. Distribution and composition of matrix protein in the radula teeth of the chiton *Acanthopleura hirtosa*. *Mar. Biol.* **109**, 281–286 (1991).
- Kuhlman, K. R., Martens, R. L., Kelly, T. F., Evans, N. D. & Miller, M. K. Fabrication of specimens of metamorphic magnetite crystals for field ion microscopy and atom probe microanalysis. *Ultramicroscopy* **89**, 169–176 (2001).
- Larson, D. J. *et al.* Analysis of bulk dielectrics with atom probe tomography. *Microsc. Microanal.* **14**, 1254–1255 (2008).
- Chen, Y., Ohkubo, T., Kodzuka, M., Morita, K. & Hono, K. Laser-assisted atom probe analysis of zirconia/spinel nanocomposite ceramics. *Scr. Mater.* **61**, 693–696 (2009).
- Marin, F., Luquet, G., Marie, B. & Medakovic, D. Molluscan shell proteins: primary structure, origin, and evolution. *Curr. Top. Dev. Biol.* **80**, 209–276 (2008).
- Law, N., Caudle, M. & Pecoraro, V. Manganese redox enzymes and model systems: properties, structures, and reactivity. *Adv. Inorg. Chem.* **46**, 305–440 (1998).
- Evans, L. A., Macey, D. J. & Webb, J. Characterization and structural organization of the organic matrix of the radula teeth of the chiton *Acanthopleura hirtosa*. *Phil. Trans. R. Soc. Lond. B* **329**, 87–96 (1990).
- Lehn, J. Supramolecular chemistry. *J. Chem. Sci.* **106**, 915–922 (1994).
- Hellman, O., Vandenbroucke, J., Rüsing, J., Isheim, D. & Seidman, D. Analysis of three-dimensional atom-probe data by the proximity histogram. *Microsc. Microanal.* **6**, 437–444 (2002).
- Addadi, L., Joester, D., Nudelman, F. & Weiner, S. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chemistry* **12**, 980–987 (2006).
- Fantner, G. *et al.* Sacrificial bonds and hidden length dissipate energy as mineralized fibrils separate during bone fracture. *Nature Mater.* **4**, 612–616 (2005).
- Bas, P., Bostel, A., Deconihout, B. & Blavette, D. A general protocol for the reconstruction of 3D atom probe data. *Appl. Surf. Sci.* **87**, 298–304 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by the Canadian National Sciences and Engineering Research Council and the US National Science Foundation (DMR-0805313). FIB, SEM and (S)TEM work was performed in the Northwestern University Atomic and Nanoscale Characterization and Experimental Center (NUANCE) supported by NSF-NSEC, NSF-MRSEC, the Keck Foundation, the State of Illinois, and Northwestern University. APT measurements were performed at the Northwestern University Center for Atom Probe Tomography (NUCAPT) supported by NSF-MRI (DMR-0420532) and ONR-DURIP (N00014-0400798, N00014-0610539, N00014-0910781). We thank A. Tayi for help in preparing chitin thin film samples; B. Myers (NUANCE) for help with FIB; D. Isheim (NUCAPT), and D. Lawrence, R. Uffig, T. Prosa and M. Greene (Cameca) for their assistance with APT; and D. Seidman for discussions.

Author Contributions L.M.G. and D.J. designed the experiments, analysed the data and prepared the manuscript. L.M.G. performed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J. (d-joester@northwestern.edu).

METHODS

Consumables. MgCl_2 , HCl , NaOH , ethanol, NaOCl (VWR); glutaraldehyde, ultra-smooth carbon adhesive tabs (Electron Microscopy Sciences); propylene oxide, Araldite 502, conductive liquid silver paint (Ted Pella); instant all-purpose brush-on cyanoacrylate adhesive (Krazy Glue); EPO-TEK H2OE conductive silver epoxy (Epoxy Technology); CarbiMet II SiC grinding paper, Metadi supreme polycrystalline aqueous diamond polishing suspension, Microcloth polishing cloths (Buehler); Chitin (TCI America); Hexafluoroisopropanol (Sigma-Aldrich); silver pellets, chromium-plated tungsten rods (Kurt J. Lesker Company). Unless otherwise noted, all solutions were prepared in ultrapure water ($\rho = 18.2 \text{ M}\Omega \text{ cm}$) prepared with a Barnstead Nanopure UF+UV ultrapure water purification system (Thermo-Fisher Scientific). Artificial seawater was prepared using Instant Ocean sea salt to a final salinity of 35 g l^{-1} .

Chitin tooth isolation and preparation. *Chaetopleura apiculata* (Say, 1834) were acquired live from the Wood's Hole Marine Biological Laboratory. Radulae were extracted by dissection after narcotization of the specimens in aqueous magnesium chloride (0.38 M , 20 min , 15°C). Radulae were fixed in 3% glutaraldehyde in artificial sea water overnight at 4°C . Samples were washed once in artificial sea water and water and then dehydrated in a graded ethanol series in water (50, 70, 90 and 100 vol%). Prior to resin infiltration the samples were transferred into propylene oxide. The samples were embedded in Araldite resin and polymerized overnight at 60°C . The embedded samples were ground using progressively finer grits of SiC grinding paper (400, 600, 800, 1,200 and 2,000 grit) and polished using polycrystalline aqueous diamond polishing suspensions (6, 3, 1.5 and $0.1 \mu\text{m}$). The polished cross-sections were secured to an aluminium stub with cyanoacrylate adhesive, coated with 25 nm of Pt by magnetron sputtering with a turbo-pumped Desk III (Denton Vacuum), then grounded to the stub with conductive liquid silver paint.

Geological magnetite samples. Geological magnetite single crystals were acquired from ABCRox and Commercial Crystal Laboratories. The ABCRox sample was cut along the (111) face with a low-speed diamond saw (Buehler Isomet) ground using progressively finer grits of SiC grinding paper (400, 600, 800, 1,200 and 2,000 grit) and polished using polycrystalline aqueous diamond polishing suspensions (6, 3, 1.5 and $0.1 \mu\text{m}$). The polished sample was secured to an aluminium stub using conductive silver epoxy. The Commercial Crystal Laboratories sample was purchased one-side polished with a (100) orientation and then secured to an aluminium stub with a conductive adhesive carbon tab.

Chitin film purification and preparation. Crude chitin was purified by adapting previously published protocols^{31–33}. The chitin was washed twice in water. The chitin was then stirred with HCl (1.5 M , 6 h , 25°C), then washed twice with water. The acid-treated chitin was further stirred with NaOH (1.5 M , 3 h , 25°C) and washed twice with water. The chitin was then treated with NaOCl (10% , 5 min , 25°C) then washed twice with water. The purified sample was then frozen in liquid nitrogen and lyophilized. The lyophilized chitin was dissolved at 1 mg ml^{-1} in hexafluoroisopropanol. The following substrate preparation and spin coating steps were carried out in a nitrogen-filled oxygen-free and water-free glove box to minimize contamination of surfaces with organic materials from the atmosphere. Silicon wafers were first coated with a 28-nm layer of Cr by thermal evaporation (MBraun) at a base pressure below 10^{-5} Pa . Deposit thickness was monitored using a quartz crystal monitor (Inficon). The thermal evaporator was vented with dry nitrogen and the coated substrates were transferred within the glove box to a spin coater (Specialty Coating Systems G3P-B) where they were spin-coated with a chitin film from the hexafluoroisopropanol solution at 3,000 r.p.m. The spin-coated substrates were then coated with 200 nm of Ag at a base pressure below 10^{-5} Pa . The wafer was secured to an aluminium stub with a conductive adhesive carbon tab.

TEM sample preparation. TEM lamellae were prepared from a polished cross-section of the chiton tooth following established procedures with a DualBeam scanning electron microscope and FIB instrument (Helios NanoLab, FEI)³⁴. First, a strap of platinum (FIB-Pt) was deposited over a region of interest on the polished cross-section using the ion beam (30 kV , 93 pA) to decompose locally an organometallic precursor gas (methyl cyclopentadienyl trimethyl platinum $[\text{C}_5\text{H}_5\text{Pt}(\text{CH}_3)_3]$). Two trenches were then milled out (30 kV , 6.5 nA) on either side of a $2\text{-}\mu\text{m}$ -thick slice of material. The slice of material was cut free (30 kV , 2.8 nA) from the substrate on three sides, leaving only a small connecting bridge. An *in situ* tungsten nanomanipulator probe (Omniprobe) was attached to the free side of the substrate using FIB-Pt (30 kV , 93 pA). The remaining connection to the substrate was milled away (30 kV , 93 pA) and the probe was retracted with the sample. The sample was then welded to a copper TEM half-grid (Omniprobe) using FIB-Pt and the connection to the probe was milled away (30 kV , 93 pA). The lamella was successively thinned to $\sim 100 \text{ nm}$ at 30 kV (93 pA) at a $1\text{--}2^\circ$ incidence angle grazing milling condition. The majority of the surface amorphization and gallium implantation was removed by low-angle milling ($\sim 7^\circ$) at 5 kV and 2 kV (28 pA) to thin the sample to approximately $60\text{--}80 \text{ nm}$.

SEM/TEM methods. Elemental mapping of polished cross-sections was performed using EDS in the Dual-Beam SEM/FIB operating at 15 kV using a Quantax X-flash silicon drift detector (Bruker AXS). Selected area electron diffraction patterns were acquired using a Hitachi H-8100 operating at 200 kV . High-angle annular dark-field STEM images and electron energy-loss spectra were acquired using a JEOL 2100F operating at 200 kV equipped with a Gatan electron energy loss spectrometer. Electron energy-loss spectroscopy maps were acquired with a dwell time of 1 s per pixel, a probe convergence angle of 10 mrad and a collection angle of 50 mrad . Pre-edge backgrounds were fitted using a power-law function.

APT sample preparation. Samples for APT were prepared using the SEM/FIB instrument using established protocols^{35–37}. First, a layer of platinum was deposited over a region of interest ($2 \mu\text{m} \times 25 \mu\text{m}$) on the polished cross-section using the ion beam (30 kV , 93 pA). For the geological samples where no sputtered or evaporated metal coating was present a protective 100 nm layer of platinum was deposited first using the electron beam (5 kV , 5.5 nA) before the ion beam was turned on. A wedge of material below the Pt strap was cut out on three sides. The wedge was attached to an *in situ* nano-manipulator (Omniprobe) using FIB-deposited Pt before cutting the final edge free. Segments $1\text{--}2\text{-}\mu\text{m}$ wide were cut from the wedge and sequentially affixed to the tops of Si posts in an array (Cameca Scientific Instruments) with FIB-Pt. Each tip was shaped and sharpened using annular milling patterns of increasingly smaller inner and outer diameters³⁵. The majority of the amorphized surface region and implanted gallium ions in the tip surface was removed by milling at 5 kV and 2 kV , at 89 pA .

APT methods. APT analyses were conducted in a Cameca local-electrode APT (LEAP 4000XSi) using a pulsed laser ($\lambda = 355 \text{ nm}$, $160\text{--}500 \text{ kHz}$, $40\text{--}75 \text{ pJ}$ per pulse). The direct-current potential on a microtip during APT was controlled to maintain an evaporation rate of either 0.0025 or 0.005 ions per laser pulse. The base temperature of the microtip was maintained at 40 K or 60 K and the ambient vacuum pressure was $< 10^{-8} \text{ Pa}$.

Three-dimensional reconstruction of APT data was performed using the IVAS atom-probe data visualization and analysis software (Cameca) using published algorithms, assuming a hemispherical tip shape^{30,38}. Standard reconstruction parameters were used with an electric-field-dependent radius (r). The average evaporation field (F_e) of the biogenic chiton tooth magnetite (23.5 V nm^{-1}) was measured by acquiring either SEM or STEM images of microtips after APT analysis. Atomic volumes of Fe and O for the reconstruction were calculated from published values of ionic radii and the known magnetite crystal structure^{39,40}. The atomic volume assigned to oxygen was increased to account for the oxygen deficiency in the measured stoichiometry. Proximity histograms²⁷ were calculated using a carbon isodensity surface (Supplementary Fig. 4) of one carbon atom per cubic nanometre and then extrapolating to the fibre axis.

APT three-dimensional renderings. Three-dimensional renderings of APT data were created in Cameca Integrated Visualization and Analysis Software (IVAS, www.cameca.com/support/ivas.aspx). The carbon containing ions (C^{2+} , CO_2^+ , CNO^+) were only rendered where the local density exceeded 0.75 nm^{-3} , owing to the spectral overlap between CO_2^+ and $^{56}\text{FeO}_2^{2+}$ in the magnetite. CO^+ was not rendered, owing to the overlap with $^{56}\text{Fe}^{2+}$. For clarity, only $5\text{--}10\%$ of the background Fe and O ions were rendered.

APT two-dimensional concentration plots. Two-dimensional concentration plots were created with a pixel size of 0.1 nm . The three-dimensional grid voxel size was $0.2 \text{ nm} \times 0.2 \text{ nm} \times 0.2 \text{ nm}$ and the delocalization parameters were 2.5 nm in x and y and 1.25 nm in z . The edge of the field of view (where the ions from the tip hit the local electrode) is marked with a white dashed line. The concentrations were averaged over a 3-nm -thick region.

1. Tsai, G. J., Su, W. H., Chen, H. C. & Pan, C. L. Antimicrobial activity of shrimp chitin and chitosan from different treatments and applications of fish preservation. *Fish. Sci.* **68**, 170–177 (2002).
2. Li, J., Revol, J. F. & Marchessault, R. H. Effect of degree of deacetylation of chitin on the properties of chitin crystallites. *J. Appl. Polym. Sci.* **65**, 373–380 (1997).
3. Sikorski, P., Hori, R. & Wada, M. Revisit of chitin crystal structure using high resolution X-ray diffraction data. *Biomacromolecules* **10**, 1100–1105 (2009).
4. Giannuzzi, L. A. & Stevie, F. A. A review of focused ion beam milling techniques for TEM specimen preparation. *Micron* **30**, 197–204 (1999).
5. Miller, M. K., Russell, K. F. & Thompson, G. B. Strategies for fabricating atom probe specimens with a dual beam FIB. *Ultramicroscopy* **102**, 287–298 (2005).
6. Thompson, K. F. *et al.* *In situ* site-specific specimen preparation for atom probe tomography. *Ultramicroscopy* **107**, 131–139 (2007).
7. Miller, M. K., Russell, K. F., Thompson, K., Alvis, R. & Larson, D. J. Review of atom probe FIB-based specimen preparation methods. *Microsc. Microanal.* **13**, 428–436 (2007).
8. Miller, M. K. *Atom Probe Tomography: Analysis at the Atomic Level* 139–153 (Plenum, 2000).
9. Shannon, R. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **32**, 751–767 (1976).
10. Fleet, M. The structure of magnetite. *Acta Crystallogr. B* **37**, 917–920 (1981).

Contrasting crustal production and rapid mantle transitions beneath back-arc ridges

Robert A. Dunn¹ & Fernando Martinez²

The opening of back-arc basins behind subduction zones progresses from initial rifting near the volcanic arc to seafloor spreading¹. During this process, the spreading ridge and the volcanic arc separate and lavas erupted at the ridge are predicted to evolve away from being heavily subduction influenced (with high volatile contents derived from the subducting plate)^{2–5}. Current models^{4,6–8} predict gradational, rather than abrupt, changes in the crust formed along the ridge as the inferred broad melting region beneath it migrates away from heavily subduction-influenced mantle. In contrast, here we show that across-strike and along-strike changes in crustal properties at the Eastern Lau spreading centre are large and abrupt, implying correspondingly large discontinuities in the nature of the mantle supplying melt to the ridge axes. With incremental separation of the ridge axis from the volcanic front of as little as 5 km, seafloor morphology changes from shallower complex volcanic landforms to deeper flat sea floor dominated by linear abyssal hills, upper crustal seismic velocities abruptly increase by over 20%, and gravity anomalies and isostasy indicate crustal thinning of more than 1.9 km. We infer that the abrupt changes in crustal properties reflect rapid evolution of the mantle entrained by the ridge, such that stable, broad triangular upwelling regions, as inferred for mid-ocean ridges^{9,10}, cannot form near the mantle wedge corner. Instead, the observations imply a dynamic process in which the ridge upwelling zone preferentially captures water-rich low-viscosity mantle when it is near the arc. As the ridge moves away from the arc, a tipping point is reached at which that material is rapidly released from the upwelling zone, resulting in rapid changes in the character of the crust formed at the ridge.

The Lau back-arc basin is a triangular-shaped extensional basin bordered by the Lau Ridge remnant arc to the west and the active Tofua arc to the east^{1,2,11}. Within the basin, the 400-km-long Eastern Lau spreading centre (ELSC) initially formed by propagating southward with its tip near the arc volcanic front¹². Magnetic data¹³ show that the ELSC is spreading nearly symmetrically but with decreasing total opening rates from north to south (96 mm yr^{−1} to 39 mm yr^{−1}); the northern end of the ELSC is presently more than twice as far from the Tofua arc than is the southern end (103 km versus 40 km). The present axis of the ELSC is a key region over which the mantle source composition changes from mid-ocean-ridge basalt (MORB)-like at the Central Lau spreading centre to the north to arc-like at the Valu Fa ridge to the south³. This and other systematic variations in axial crustal properties, such as ridge morphology¹⁴ and seismic structure¹⁵, correlate with the changing proximity of the ridge to the arc volcanic front^{3,4}. The observations are consistent with a decreasing 'subduction influence' in the mantle as the ridge shifts away from the arc. Transitions in along-axis crustal properties are known to occur near ridge-axis offsets, which discontinuously shift the ridge segments relative to the arc front. However, the across-axis evolution of the crust, which reveals the continuous history of changing mantle influence as each section of ridge has moved further from the arc, has not been investigated.

We present results of seismic tomography (Fig. 1) and geophysical studies (Fig. 2) of the ELSC axis and flanks spanning two complete

ridge segments and portions of the adjoining segments. In plan view, the upper crust formed on the ELSC is characterized by distinct zones of relatively low, intermediate and high compressional wave velocities (Fig. 1b), which are most probably related to variations in both bulk-rock porosity and composition¹⁵. The velocity variations spatially correlate with major-element changes in seafloor lavas (Fig. 1c), geochemical estimates of changes in source volatile content (and other arc-related components), and with changes in the vesicularity of seafloor samples^{2,3,8,16}. Therefore we suggest that the seismic structure, via the effects of crustal porosity and composition, provides a record of changes in the initial volatile content of melts fed to the ridge.

Changes in upper crustal seismic velocity (Fig. 1b) closely mirror changes in seafloor morphology and depth (Fig. 2a) and gravity (Fig. 2b), and imply abrupt changes in crustal production controlled by the volatile (H₂O) content of the mantle¹⁷. Together with geochemical data^{2,3,8,16,18}, these observations delineate two distinct crustal domains separated by a variable transition region. In the spreading direction, observed step-like transitions between these domains suggest that abrupt changes in mantle source composition take place over as little as 5 km of incremental spreading.

In Fig. 2, the older crust into which the ELSC propagated is labelled as Domain I-type crust. Crust formed at the ELSC is labelled Domains II and III (Figs 1–2). Domain II is composed of crust that was produced soon after the ridge propagated southward near the arc volcanic front. Here, shallow and complex volcanic terrains imply high mantle heterogeneity and enhanced crustal production. The terrain is dominated by short arcuate segments with central highs and distal lows that are concave towards the present ridge axis. Domain II terrains are associated with low seismic velocities (~ 4.0 km s^{−1} at 1 km depth), implying high porosities¹⁹ (>15%) and arc-related mineralogies¹⁵. These velocities are similar to velocities previously detected on the Valu Fa ridge to the south^{15,20} and are anomalously slow compared to mid-ocean ridges, which are typically 5–6 km s^{−1} for this depth (Fig. 1 inset). The ~ 500 -m shallower bathymetry and low gravity values suggest that the crust is on average about 1.9 km thicker than crust produced closer to the ELSC in more recent times. The few rock samples in these off-axis areas range from basalts to andesites^{2,8,18}. The low velocities, geochemistry and thick crust indicate that during this period the ridge upwelling zone beneath the spreading centre entrained a significant degree of slab-derived water. The addition of water to the mantle is expected to increase the degree of melting by lowering the mantle solidus^{5,17,21}, resulting in a thicker crust with more vesicular andesitic lavas². A current morphologic analogue of Domain II terrains may be found along the Fonualei spreading centre in the northeastern Lau basin. The Valu Fa ridge may also be an active analogue, though it exhibits less variable morphology, suggesting relatively less heterogeneity in the mantle source.

Domain III crust is characterized by sea floor ~ 500 m deeper (at $\sim 2,600$ m depth) and linear abyssal hill tectonic fabric typical of intermediate-spreading-rate mid-ocean ridges (Fig. 2a). Domain III spatially correlates with basaltic lavas⁸ and is further characterized by upper crustal velocities (on average 5 km s^{−1} at 1 km depth) that are closer to typical oceanic upper crustal velocities (Fig. 1 inset); estimated

¹Department of Geology and Geophysics, University of Hawaii, Honolulu, Hawaii 96822, USA. ²Hawaii Institute of Geophysics and Planetology, University of Hawaii, Honolulu, Hawaii 96822, USA.

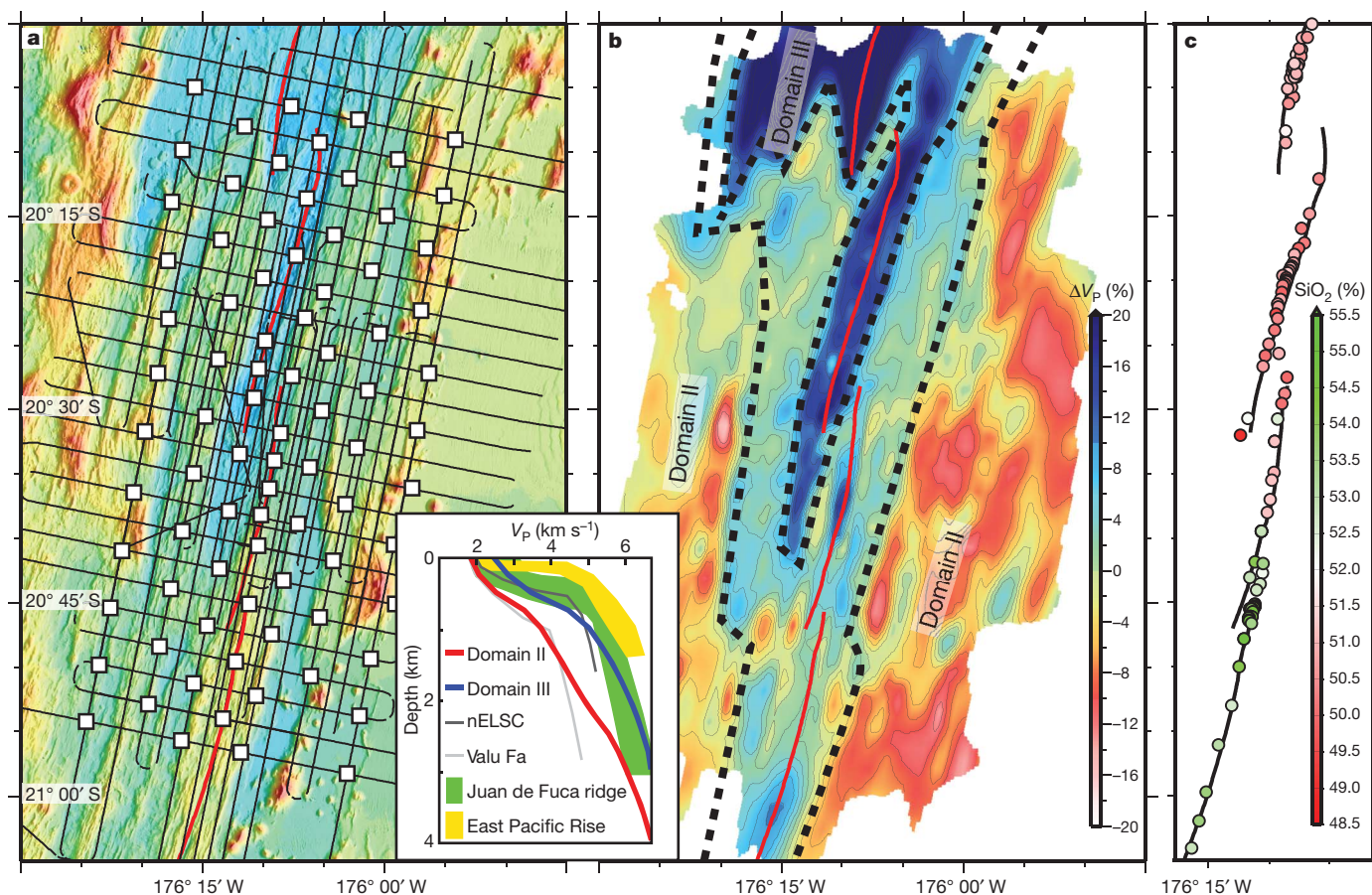


Figure 1 | Layout of the seismic experiment and map view of compressional wave velocity variations in the upper crust. **a**, Square symbols indicate the positions of 83 ocean-bottom seismometers that recorded P-wave arrivals from 8,763 airgun shots that occurred along the black lines. **b**, A map-view tomographic image of the P-wave velocity (V_p) structure at an average depth of 1 km beneath the seafloor. Dotted lines indicate the boundaries of the different crustal domains as described in Fig. 2 and the text. Velocities are given as a percentage variation from the average (4 km s^{-1}). Inset, a comparison of average Domain II and III velocity profiles with multi-channel-seismic-derived V_p profiles from the axis of the Valu Fa ridge and the northern Eastern Lau

porosities for this domain are $<10\%$ (ref. 19). The deeper sea floor probably reflects an isostatic response to both higher rock density and thinner crust. The highest-velocity regions in the study area correspond to the deep crust formed along the northernmost spreading segment (Fig. 1b). On the basis of the seismic velocity and bathymetry maps, and the geochemical analyses of rocks collected in this area^{2,8,16}, relatively low-volatile-content melts with basaltic compositions may have formed crust along this ridge segment for the past ~ 0.39 million years (16 km of spreading). A narrow region of Domain III crust is located along the entire axis of the ELSC segment just to the south (Fig. 1b), suggesting a recent switch-over (81 thousand years ago) from a ridge segment that formerly tapped volatile-rich melts to one that now taps relatively volatile-poorer melts.

Sandwiched between the two domains is a transitional region, varying from ~ 5 – 10 km width, with intermediate-depth abyssal hill fabrics (Fig. 1a) and intermediate upper-crustal seismic velocities (Fig. 2b). Along the northern ELSC segment, the switch to transitional crust occurred over a few kilometres of spreading roughly 0.58 million years ago and a later switch to Domain III crust occurred abruptly at 0.39 million years ago. Along the central ELSC, the corresponding changes occurred at roughly 0.261 million years ago and 0.081 million years ago, respectively. Along the southern ELSC (south of the overlapping spreading centre at $20^\circ 30' \text{ S}$), the ridge axis appears to be currently in the transitional phase and seafloor rock samples with

spreading centre, along with velocities of typical fast- and intermediate-spreading-rate crust (after ref. 15). Domain II velocities are similar to those of the Valu Fa ridge and are much lower than for typical mid-ocean ridges; Domain II velocities are consistent with both a thicker seismic layer 2A and lower layer 2B velocities. **c**, Geological samples collected along the axis of the ridge exhibit major-element, trace-element, and isotopic-ratio variations that closely correlate with the geophysical observations^{8,16}. Here we show SiO_2 content (C.H. Langmuir, personal communication) as an indicator of the basalt-to-andesite transition that occurs along the ELSC.

andesitic compositions are found here⁸ (Fig. 1c). The exception is that a small northern portion of this ridge section currently appears to tap Domain III melts. Here, the switch from transitional-type crust to Domain III crust (near $20^\circ 39' \text{ S}$), as determined by the seismic image, occurs near a well-documented major-element, trace-element and isotopic geochemical boundary between more heavily subduction-influenced lavas along the ridge axis to the south to less so along the ridge axis to the north^{8,16} (Fig. 1c).

The combined data sets show significant correlation between crustal seismic velocity structure and changes in seafloor morphology, gravity and geochemistry. For a single spreading segment, abrupt transitions in crustal properties take place in a step-like fashion with only a few kilometres of incremental spreading. These abrupt transitions are unlike along-axis transitions that occur at mid-ocean ridges, which are thought to reflect dynamic threshold effects associated with changing spreading rate^{22,23} or mantle temperature²⁴. We infer that the crustal transitions at the ELSC reflect abrupt changes in the composition of the mantle supplying melt to the back-arc ridge axes. Thus, the observations cast doubt on models of arc-proximal back-arc melting regions that are similar to the broad ~ 200 -km-wide triangular regions imaged⁹ and modelled at mid-ocean ridges¹⁰.

Such broad melting regimes would be expected gradually to entrain the various mantle wedge compositional domains as they migrate from arc-proximal to arc-distal settings and produce gradational changes in

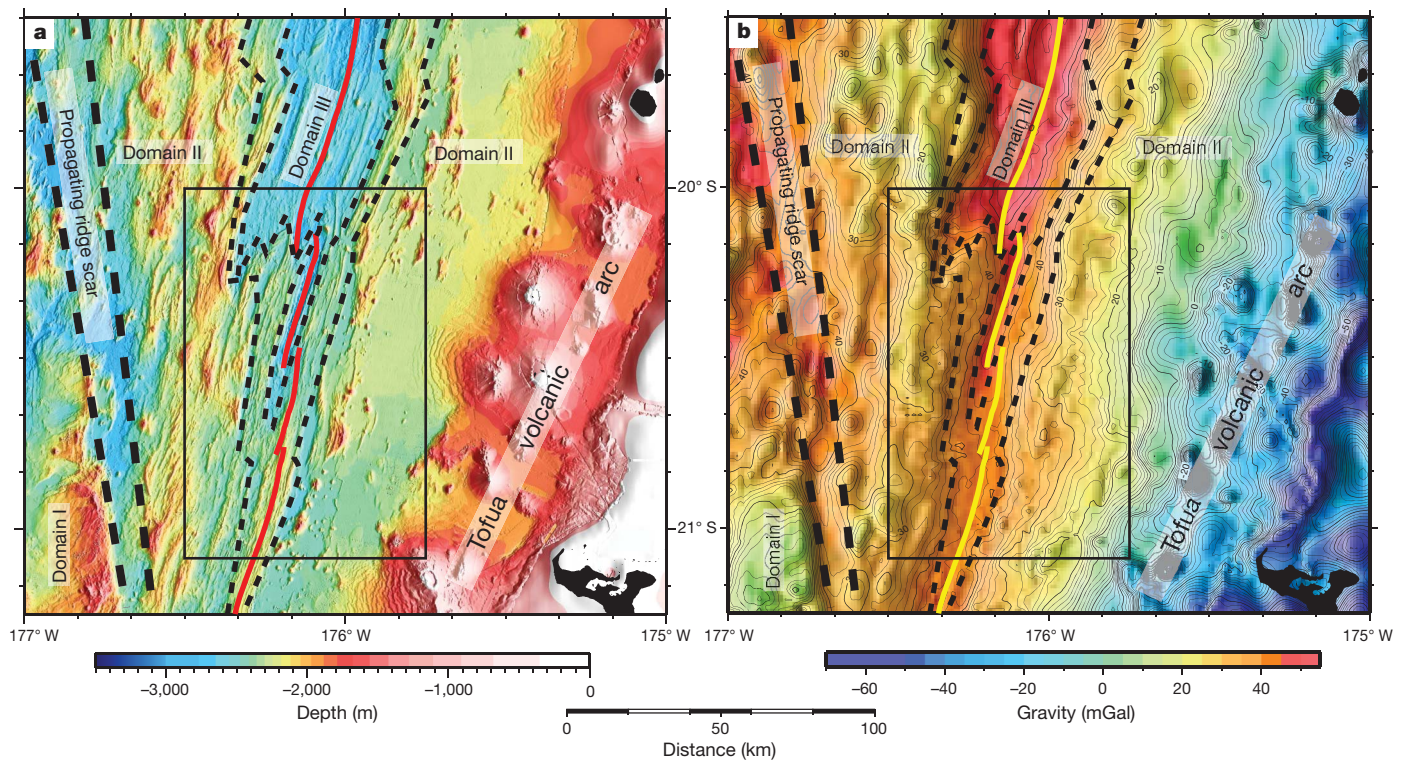


Figure 2 | Seafloor bathymetry and Bouguer gravity anomalies in the study area. Red lines show the axis of the ELSC; the area of the seismic experiment is outlined by a box. **a**, The seafloor bathymetry is characterized by an older domain of pre-existing crust (Domain I) into which the ELSC propagated and two types of ELSC terrains labelled Domains II and III, separated by a variable transitional region demarcated by the dashed lines (see text). The two terrains closely follow the domains of low and high crustal seismic velocity shown in Fig. 1b. We note that near the active Tofua volcanic arc, volcanoclastic

sediments largely cover Domain II terrains. **b**, Bouguer gravity anomalies derived from ship and satellite data (contour interval is 2 mGal). The regional gradient associated with the slab was removed as described in ref. 4. Relative gravity lows and highs correspond to Domains II and III, respectively, separated by abrupt gradients. Local isostatic balance, assuming a crustal density of $2,700 \text{ kg m}^{-3}$ and a mantle density of $3,300 \text{ kg m}^{-3}$, indicates that Domain II crust is more than 1.9 km thicker than Domain III crust.

crustal properties. At the ELSC, the step-like changes in crustal properties suggest rapid transitions in the nature of the mantle entrained in the melting regime. We infer (Fig. 3) that these changes are enabled by a viscosity contrast of one to several orders of magnitude between a hydrous mantle near the arc and a relatively drier mantle further from the volcanic front²⁵. We suggest that when near the volcanic arc, back-arc spreading centres preferentially advect hydrous, low-viscosity mantle, possibly augmented by dynamic buoyant upwelling, as inferred for the arc itself^{26,27}. This in turn produces the low-seismic-velocity, shallow, thick and volcanically complex crust observed. With increasing distance from the arc, however, a point is reached at which the connection to the low-viscosity hydrous mantle is broken and the spreading centre subsequently advects less-hydrous and higher-viscosity mantle. As the low-viscosity mantle is let go from the ridge upwelling zone, downwelling may occur on the arc side of the ridge, further suppressing melting of the higher-water-content mantle. At the ridge axis, mid-ocean-ridge-like crust is thereafter formed, although such crust still contains higher water content than at typical mid-ocean ridges and has trace-element enrichment and depletion characteristic of subduction, which can persist for hundreds of kilometres behind the arc volcanic front, giving back-arc basin basalts their distinctive character^{3,5,8}.

We hypothesize (Fig. 3) that geochemically and rheologically contrasting mantle domains with abrupt boundaries exist in arc-proximal regions of the mantle wedge, rather than gradational variations, such that individual spreading centres in this region can abruptly switch from advecting one to the other domain with only a few kilometres of incremental spreading. Even neighbouring ridge segments, separated by small non-transform discontinuities, can erupt lavas derived from contrasting mantle domains for long periods of time. For example, the juxtaposition of distinct domains across the northern overlapping

spreading centre is estimated to have lasted for roughly 0.31 million years. One can even see the duelling propagation ‘scars’ in the depth contrast (Fig. 2a) and velocity structure (Fig. 1b) as the W-shaped boundary that reflects the contrasting mantle domains and crust spread by the two neighbouring ridge segments. For an 8-km ridge offset to have acted as a persistent boundary between mantle domains implies either a sharp chemical distinction between the mantle regions that fed the two ridge axes or a rapidly changing mantle flow field, or both.

METHODS SUMMARY

The L-SCAN experiment was carried out in January–March 2009, using the R/V *Langseth* and its 36-element, 6,600-cubic-inch airgun array towed at 9.0 m depth. The seismic data were recorded on 83 ocean-bottom seismometers (58 instruments, 25 of which were deployed twice) obtained via the National Ocean Bottom Seismograph Instrument Pool (<https://obsip.ucsd.edu>). Each unit contained a three-component seismometer and a hydrophone; the sample interval was 5 ms. Shots occurred every 450–500 m along 57 refraction lines. The ocean bottom seismometers were located using the travel times of the direct water wave for shots within 3 km of an instrument and a Bayesian grid search algorithm²⁸; the average 1σ location uncertainty was 6 m.

At about 2 km to 12 km from an instrument, the first wave energy to arrive is a refraction from the upper crust (layer 2) with turning depths of 250 m to 2 km. Layer 3 refractions occur at greater distances (out to ~30 km). All layer 2 arrivals and layer 3 arrivals to ~20 km offset were used in this study. The arrivals are high quality and could be picked to within 1–3 samples; including all estimated experimental errors, the root-mean-square uncertainty of all travel times is 15 ms. The compressional (P)-wave velocity image was constructed using a nonlinear iterative tomographic technique²⁹ that modelled 139,986 travel-time data points for three-dimensional P-wave velocity structure and depth-varying anisotropy. Ray tracing was accomplished via a shortest path technique followed by a ray optimization procedure³⁰. Although the number of data exceeds the number of unknowns by a

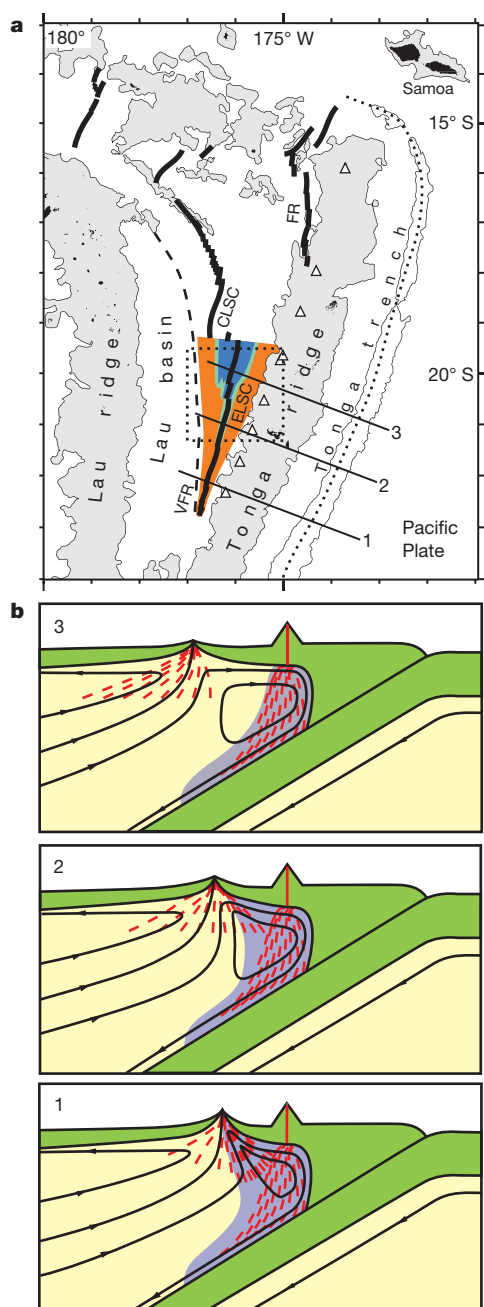


Figure 3 | Map of the Lau basin and cartoon interpretations of the formation of the crustal domains via changes in mantle wedge composition and flow patterns. **a**, Map of the Lau basin showing simplified interpretations of the velocity domains shown in Figs 1 and 2 extrapolated to the Valu Fa Ridge (VFR) (Domain II in orange, Domain III in blue, transition zone in green). The dotted box shows the area of Fig. 2a. The numbered lines show the locations of the vertical sections depicted in the panels of **b**. The dashed line is the boundary formed by the southward propagation of the ELSC. CLSC, Central Lau spreading centre. FR, Fonualei rift. **b**, (1) Formation of Domain II crust. Dynamic upwelling of the mantle occurs beneath the spreading axis when it is near the volcanic front and entrains low-viscosity mantle hydrated by water released from the subducting slab. Crust formed by the spreading centre in this position is arc-like, heterogeneous, thick, and has high porosity. (2) Formation of transitional crust. As the spreading centre separates from the volcanic front, the connection to the hydrous and low-viscosity mantle is abruptly cut off. Melt generation transitions to that of passive upwelling driven by plate separation at the ridge axis. Thinner and lower-porosity crust is formed at the spreading centre. (3) Formation of Domain III crust. At greater separation from the volcanic front, melt generation at the ridge results from passive upwelling and is separated from the hydrous and dynamic mantle upwelling beneath the arc. Mantle melted beneath the arc partly recirculates through the ridge system and the crust produced at the ridge has a lower porosity owing to lower slab water input, is thinner, and is depleted owing to prior melt extraction. The red dashes in panels 1–3 depict melt flow, which may be independent of solid mantle flow indicated by schematic streamlines. The blue area indicates hydrous mantle, the yellow area is nominally anhydrous mantle and green areas are lithospheric plates.

factor of 2.5, the inversion was nevertheless regularized with weak smoothing and damping constraints. Seafloor bathymetry was included in the modelling procedure and its proper usage was extensively checked. (See also the Supplementary Information.)

Received 16 August; accepted 18 November 2010.

- Karig, D. E. Ridges and basins of the Tonga-Kermadec island arc system. *J. Geophys. Res.* **75**, 239–254 (1970).
- Hawkins, J. W. in *Backarc Basins: Tectonics and Magmatism* (ed. Taylor, B.) 63–138 (Plenum, 1995).
- Pearce, J. A. et al. in *Volcanism Associated with Extension at Consuming Plate Margins* (ed. Smillie, J. L.) *Geol. Soc. Spec. Publ.* **81**, 53–75 (1995).
- Martinez, F. & Taylor, B. Mantle wedge control on back-arc crustal accretion. *Nature* **416**, 417–420 (2002).
- Taylor, B. & Martinez, F. Back-arc basin basalt systematics. *Earth Planet. Sci. Lett.* **210**, 481–497 (2003).
- Kincaid, C. & Hall, P. S. Role of back arc spreading in circulation and melting at subduction zones. *J. Geophys. Res.* **108**, 2240, doi:10.1029/2001JB001174 (2003).
- Pearce, J. A., Stern, R. J., Bloomer, S. H. & Fryer, P. Geochemical mapping of the Mariana arc-basin system: implications for the nature and distribution of subduction components. *Geochem. Geophys. Geosyst.* **6**, doi:10.1029/2004GC000895 (2005).
- Langmuir, C. H., Bezos, A., Escrig, S. & Parman, S. W. in *Back-Arc Spreading Systems: Geological, Biological, Chemical, and Physical Interactions* (eds Christie, D. M., Fisher, C. R., Lee, S.-M. & Givens, S.) *Geophys. Monogr. AGU* **166**, 87–146 (2006).
- Forsyth, D. W. et al. Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment. *Science* **280**, 1215–1218 (1998).
- Conder, J. A., Forsyth, D. W. & Parmentier, E. M. Asthenospheric flow and asymmetry of the East Pacific Rise, MELT area. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000807 (2002).
- Gill, J. B. Composition and age of Lau Basin and Ridge volcanic rocks: implications for evolution of an interarc basin and remnant arc. *Geol. Soc. Am. Bull.* **87**, 1384–1395 (1976).
- Parson, L. M., Pearce, J. A., Murton, B. J., Hodkinson, R. A., & the RRS Charles Darwin Scientific Party. Role of ridge jumps and ridge propagation in the tectonic evolution of the Lau back-arc basin, southwest Pacific. *Geology* **18**, 470–473 (1990).
- Zellmer, K. E. & Taylor, B. A three-plate kinematic model for Lau Basin opening. *Geochem. Geophys. Geosyst.* **2**, doi:10.1029/2000GC000106 (2001).
- Martinez, F., Taylor, B., Baker, E. T., Resing, J. A. & Walker, S. L. Opposing trends in crustal thickness and spreading rate along the back-arc Eastern Lau Spreading Center: implications for controls on ridge morphology, faulting, and hydrothermal activity. *Earth Planet. Sci. Lett.* **245**, 655–672 (2006).
- Jacobs, A. M., Harding, A. J. & Kent, G. M. Axial crustal structure of the Lau back-arc basin from velocity modeling of multichannel seismic data. *Earth Planet. Sci. Lett.* **259**, 239–255 (2007).
- Escrig, S., Goldstein, S. L., Langmuir, C. H. & Michael, P. J. Mantle source variations beneath the Eastern Lau Spreading Center and the nature of subduction components in the Lau basin–Tonga arc system. *Geochem. Geophys. Geosyst.* **10**, doi:10.1029/2008GC002281 (2009).
- Stolper, E. & Newman, S. The role of water in the petrogenesis of Mariana trough magmas. *Earth Planet. Sci. Lett.* **121**, 293–325 (1994).
- Hawkins, J. W. & Melchior, J. T. Petrology of Mariana Trough and Lau Basin basalts. *J. Geophys. Res.* **90**, 11431–11468 (1985).
- Carlson, R. L. & Herrick, C. N. Densities and porosities in the oceanic crust and their variations with depth and age. *J. Geophys. Res.* **95**, 9153–9170 (1990).
- Turner, I. M., Pierce, C. & Sihna, M. C. Seismic imaging of the axial region of the Valu Fa Ridge, Lau Basin—the accretionary processes of an intermediate back-arc spreading ridge. *Geophys. J. Int.* **138**, 495–519 (1999).
- Davies, J. H. & Bickle, M. J. A physical model for the volume and composition of melt produced by hydrous fluxing above subduction zones. *Phil. Trans. R. Soc. Lond. A* **335**, 355–364 (1991).
- Small, C. A global analysis of mid-ocean ridge axial topography. *Geophys. J. Int.* **116**, 64–84 (1994).
- Phipps Morgan, J. & Chen, Y. J. Dependence of ridge-axis morphology on magma supply and spreading rate. *Nature* **364**, 706–708 (1993).
- Ma, Y. & Cochran, J. R. Transitions in axial morphology along the southeast Indian Ridge. *J. Geophys. Res.* **101**, 15849–15866 (1996).
- Hirth, G. & Kohlstedt, D. L. in *Inside the Subduction Factory* (ed. Eiler, J.) 83–105 (American Geophysical Union, 2003).
- Marsh, B. D. Island-arc development: some observations, experiments, and speculations. *J. Geol.* **87**, 687–713 (1979).

27. Gerya, T. V. & Yuen, D. A. Rayleigh-Taylor instabilities from hydration and melting propel 'cold plumes' at subduction zones. *Earth Planet. Sci. Lett.* **212**, 47–62 (2003).
28. Dunn, R. A. & Hernandez, O. Tracking blue whales in the eastern tropical Pacific with an ocean-bottom seismometer and hydrophone array. *J. Acoust. Soc. Am.* **126**, 1084–1094 (2009).
29. Dunn, R. A., Lekic, V., Detrick, R. S. & Toomey, D. R. Three-dimensional seismic structure of the Mid-Atlantic Ridge at 35°N: focused melt supply and non-uniform plate spreading. *J. Geophys. Res.* **110**, doi:10.1029/2004JB003473 (2005).
30. Nishi, K. A three-dimensional robust seismic ray tracer for volcanic regions. *Earth Planets Space* **53**, 101–109 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the captain, crew, and science parties of the R/V *Langseth* leg MGL0903 and R/V *Kilo Moana* leg KM0804. This work was funded as part of the NSF Ridge 2000 Program by grants OCE0426428 (R.A.D.) and OCE0732536 & OCE0727138 (F.M.). We thank J. Hammer, G. Ito, J. Sinton, B. Taylor and C. Wolfe for comments on an early version of the manuscript.

Author Contributions R.A.D. carried out the seismic data collection, analysis and modelling; F.M. carried out bathymetry data collection and processing and gravity analysis. Both authors contributed to writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.A.D. (dunn@soest.hawaii.edu).

Reliability of flipper-banded penguins as indicators of climate change

Claire Saraux^{1,2,3,4}, Céline Le Bohec^{1,2,3}, Joël M. Durant³, Vincent A. Viblanc^{1,2}, Michel Gauthier-Clerc⁵, David Beaune^{1,2}, Young-Hyang Park⁶, Nigel G. Yoccoz⁷, Nils C. Stenseth^{3,8} & Yvon Le Maho^{1,2}

In 2007, the Intergovernmental Panel on Climate Change highlighted an urgent need to assess the responses of marine ecosystems to climate change¹. Because they lie in a high-latitude region, the Southern Ocean ecosystems are expected to be strongly affected by global warming. Using top predators of this highly productive ocean² (such as penguins) as integrative indicators may help us assess the impacts of climate change on marine ecosystems^{3,4}. Yet most available information on penguin population dynamics is based on the controversial use of flipper banding. Although some reports have found the effects of flipper bands to be deleterious^{5–8}, some short-term (one-year) studies have concluded otherwise^{9–11}, resulting in the continuation of extensive banding schemes and the use of data sets thus collected to predict climate impact on natural populations^{12,13}. Here we show that banding of free-ranging king penguins (*Aptenodytes patagonicus*) impairs both survival and reproduction, ultimately affecting population growth rate. Over the course of a 10-year longitudinal study, banded birds produced 39% fewer chicks and had a survival rate 16% lower than non-banded birds, demonstrating a massive long-term impact of banding and thus refuting the assumption that birds will ultimately adapt to being banded^{6,12}. Indeed, banded birds still arrived later for breeding at the study site and had longer foraging trips even after 10 years. One of our major findings is that responses of flipper-banded penguins to climate variability (that is, changes in sea surface temperature and in the Southern Oscillation index) differ from those of non-banded birds. We show that only long-term investigations may allow an evaluation of the impact of flipper bands and that every major life-history trait can be affected, calling into question the banding schemes still going on. In addition, our understanding of the effects of climate change on marine ecosystems based on flipper-band data should be reconsidered.

The effects of climate forcing on primary and secondary production of the short austral food webs may be integrated at higher levels^{14,15}, and thus amplified in top-level predators such as seabirds. This has led to a strong interest in studying Antarctic and sub-Antarctic top predators (especially penguins, which are major consumers of the Southern Ocean ecosystem) as sensitive indicators of environmental changes^{3,4}. To understand how variability in marine resources affects their demography over the timescale of years, simultaneous investigations of variation in breeding success and survival are necessary and require long-term individual monitoring at the population scale.

Most of our present knowledge on the population dynamics of penguins is based on large flipper-banding schemes. The key advantage is that bands can be identified from a distance, avoiding recapture stress for the birds. In the 1970s, however, bands were observed, both in zoos and in the wild, to injure flipper tissues severely⁵, especially during the moult. Although many research programmes consequently abandoned banding as a precaution in the late 1980s, massive banding

schemes still continued (see references in ref. 5). Yet, as penguins power their swimming exclusively with their flippers, there has been an increasing concern about the hydrodynamic drag effect that may be induced by flipper bands (for example a 24% increase in the energy cost of swimming in captive Adélie penguins⁵). The question was then whether penguins may compensate for such effects^{10,11} and whether the impact of flipper bands would be limited in time. Although it had been assumed that the effect of flipper bands lasted for a year at most¹² (until the bird got used to the band), the question remained to be addressed in the long term. In this context, medium-term studies revealed lower breeding success and survival in Adélie penguins⁸ and a reduced breeding success in king penguins⁷. However, those pioneering findings did not result in the cessation of ongoing banding schemes. Whether or not flipper bands have a deleterious impact in the long term is, nonetheless, a crucial issue, for “it raises practical and larger ethical questions about costs and benefits of procedures in field studies”¹⁶. In addition to possibly harming penguin populations already under threat (such as penguins rehabilitated after oil spills; see references in ref. 7), the potentially negative effects of banding on demographic parameters may introduce a bias, which in turn might jeopardize any attempt to use data from banded birds to assess the impact of climate¹² on population dynamics and to predict the future

Table 1 | Observed differences between life-history traits of banded and non-banded king penguins

| Parameter | Non-banded | Banded | <i>P</i> (banding) |
|---|----------------------|----------------------|-----------------------|
| Early group | | | |
| Arrival dates | 21 Nov. ± 2 d (189) | 7 Dec. ± 3 d (167) | <0.001 |
| Breeding propensity | 0.95 (189) | 0.87 (167) | 0.04 |
| Laying dates | 29 Nov. ± 1 d (160) | 6 Dec. ± 1 d (122) | <0.001 |
| Breeding success | 0.44 (160) | 0.32 (122) | 0.05 |
| Foraging trips | 11.60 ± 0.20 d (512) | 12.70 ± 0.20 d (344) | <0.001 |
| All birds | | | |
| Overall survival over the decade | 0.36 | 0.20 | 0.04 |
| Overall/annual survival over the first period (4.5 years) | 0.62/0.90 | 0.32/0.78 | 0.01 |
| Overall/annual survival over the last period (5.5 years) | 0.57/0.90 | 0.60/0.91 | 0.82 |

Significant results are indicated in bold. Data shows mean ± s.e. The number of events (*N*) is shown in parentheses. Differences in *N* come from differences in studied stages (for example, not every bird arriving at the colony bred). Overall survival corresponds to the number of studied birds present at the colony at the end of the period divided by the number of studied birds present at the colony at the beginning of the period. Breeding propensity corresponds to the proportion of live birds that engaged in reproduction over the ten breeding seasons (that is, the number of reproduction events divided by the sum over the years of live birds). The early group is the group of birds that failed or did not engage in previous reproduction and were free to arrive early in the summer. This group is the one that most affects overall population reproductive success. For breeding analyses, *P* is the *P* value of the banding effect in the mixed model $Y = \text{banding} + (1 | \text{individual})$. For survival (investigated through schemes of presence/absence at the colony), *P* is the *P* value obtained from a Cox proportional hazards model with right censoring. Durations of foraging trips were standardized per period and then pooled together to run a single mixed-model analysis.

¹Université de Strasbourg, Institut Pluridisciplinaire Hubert Curien, 23 rue Becquerel, 67087 Strasbourg, France. ²CNRS, UMR7178, 67037 Strasbourg, France. ³Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, PO Box 1066, Blindern, N-0316 Oslo, Norway. ⁴AgroParisTech, ENGREF, 19 avenue du Maine, F-75732 Paris, France. ⁵Centre de Recherche de la Tour du Valat, Le Sambuc, 13200 Arles, France. ⁶Département Milieux et Peuplements Aquatiques, USM 0402/LOCEAN, Muséum National d'Histoire Naturelle, 75231 Paris, France. ⁷Department of Arctic and Marine Biology, University of Tromsø, N-9037 Tromsø, Norway. ⁸Institute of Marine Research, Flødevigen Marine Research Station, N-4817 His, Norway.

of breeding colonies according to scenarios of the Intergovernmental Panel on Climate Change¹³.

Using a decade-long data set, we analysed differences in reproduction and survival between two groups of electronically monitored king penguins (50 banded and 50 non-banded) breeding on Possession Island in the sub-Antarctic Crozet archipelago. These differences were investigated in relation to the birds' time of arrival for breeding and the duration of their foraging trips. Furthermore, we explored whether and how climate variability might differently affect banded and non-banded penguins.

First and foremost, our study underlines a strong negative impact of flipper banding on adult survival, to which population growth rate is most sensitive in long-lived species^{17,18} such as king penguins. The average annual survival was 5% lower in banded penguins, leading to an overall 16%-lower survival over a decade (Table 1 and Fig. 1a). Yet a breakpoint analysis revealed that the difference between the cumulative survival of banded penguins and that of non-banded penguins had one breakpoint, after 4.5 years (54 ± 3 months; Fig. 1b). Before that the mortality was 30% higher for banded birds, whereas afterwards the survival rates of the remaining banded and non-banded birds were not significantly different (Table 1 and Fig. 1). Importantly, those birds that died during the first period (banded or not) had a lower breeding success than those that survived (0.22 versus 0.30, $P = 0.047$) and the annual survival rate of banded penguins increased between the two periods (from 78% to 91%, Wilcoxon rank-sum test, $P = 0.05$). This suggests that banding has a stronger deleterious effect on low-quality birds and thus selects for high-quality individuals.

Second, we show that over the decade banded birds had significantly lower breeding propensity and success than non-banded birds (breeding probability of 0.87 versus 0.95 and breeding success of 0.32 versus 0.44; Table 1 and Supplementary Fig. 1). This could be explained by the significantly later arrival of banded birds to breed (average delay relative to non-banded birds was 16 days; Table 1). Indeed, breeding propensity was negatively affected by arrival dates, meaning that those birds arriving later were less likely to engage in breeding attempts (model 9, $P < 0.001$; Supplementary Table 1). According to life-history trade-offs¹⁸ between reproduction and survival for long-lived species, late-arriving birds may delay reproduction to the following year to avoid breeding costs that may jeopardize their own survival^{19,20}. Furthermore, in agreement with previous studies^{21,22} we found that the later in the season king penguins arrived to breed, the lower was their breeding success (Supplementary Fig. 2 and model 15, $P < 0.001$; Supplementary Table 1). This suggests an unfavourable timing in chick rearing, which then begins at the end of summer, when prey availability is much lower²³. The reduced breeding success of banded birds could also be explained by the greater durations of their foraging trips at all summer stages of the breeding cycle (Table 1 and Supplementary Fig. 3). Indeed, a longer trip may either lead to breeding failure (that is, no longer being able to sustain the fasting period energetically, the partner abandons the egg before mate relief²⁴) or directly jeopardize the survival of chicks waiting for food. Accordingly, birds failing in reproduction made significantly longer trips at sea than successful birds (21.8 versus 16.1 days, $P < 0.001$, for incubation birds and 11.5 versus 8.1 days, $P < 0.001$, for birds with chicks). Thus, the longer trip duration of the banded birds suggests that the detrimental effect of the bands can be explained by a reduced swimming and/or foraging efficiency resulting from the effect of flipper-band drag on the hydrodynamics of the bird, such as for Adélie penguins⁵.

Notably, our data clearly do not accord with the assumption that flipper bands have an impact essentially restricted to the first year after banding, which is the time suggested for birds to habituate to the handicap^{6,12}. Indeed, flipper bands also had a deleterious effect during the second half of our study (for example, $P = 0.008$ for arrival dates). These long-term effects indicate that there is no habituation to the handicap. We conclude that flipper bands lead to delayed breeding

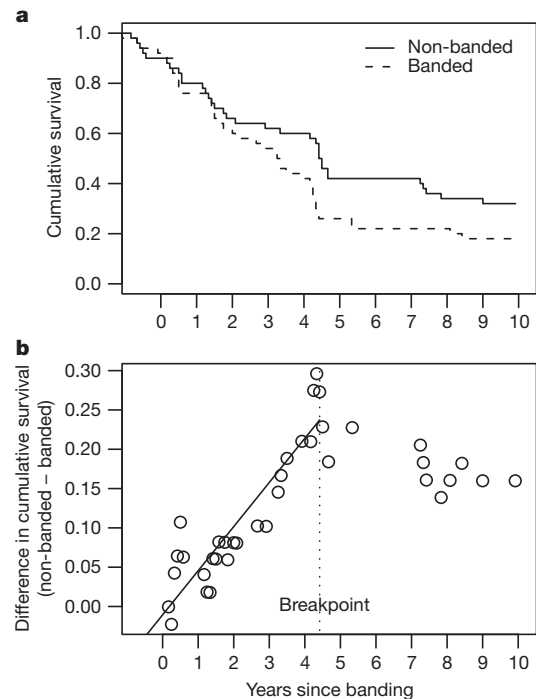


Figure 1 | Survival of banded and non-banded king penguins during the 10-year study period. **a**, Cumulative survival was lower for banded birds (dashed line) than for non-banded birds (solid line) (Cox proportional hazard model, $P = 0.04$; assumption of proportional hazards verified, $P = 0.83$). **b**, Difference between the cumulative survivals of banded and non-banded penguins over time. There is a breakpoint at 54 months (4.5 years) and the linear trend is indicated. Differences between banded and non-banded birds tended to disappear after the first 4.5 years.

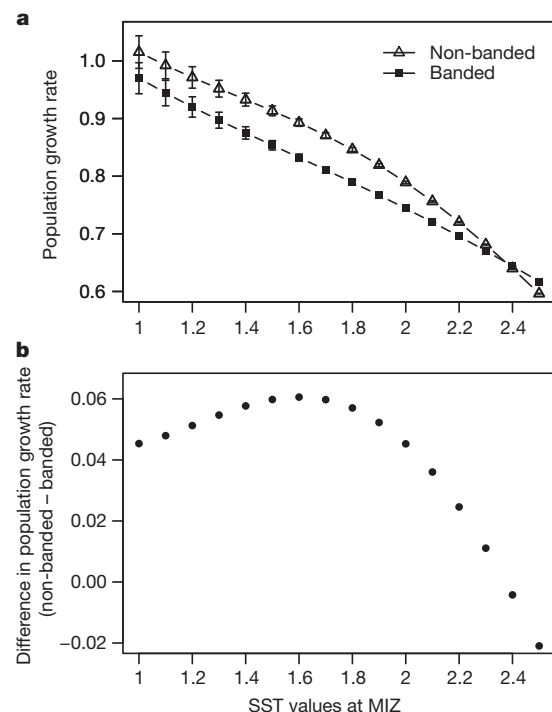


Figure 2 | Simulated population growth rates of banded and non-banded penguins as functions of SST. **a**, Growth rates of both populations according to SST at the marginal ice zone (MIZ). Error bars, s.e.m. **b**, Difference between the two growth rates. A quadratic relation well approximated the difference (Growth rate $\approx (0.27 \pm 0.01)\text{SST} - (0.09 \pm 0.00)\text{SST}^2$, $P < 0.001$ for both SST and SST^2).

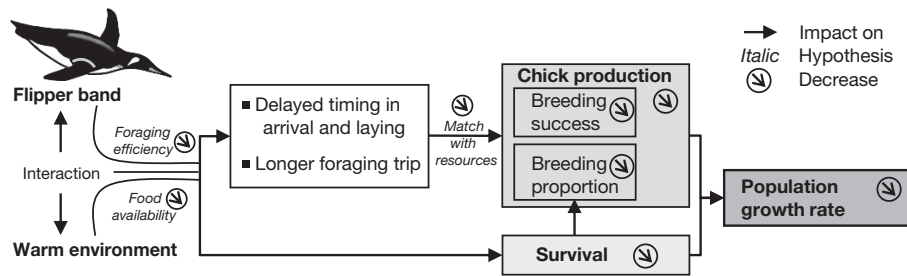


Figure 3 | Potential mechanisms involved in negative impacts of flipper bands on life-history traits and population dynamics in king penguins. Flipper bands and climate interact to affect chick production negatively, mostly through delayed timing, survival and, ultimately, population growth rate.

attempts, lower breeding propensities and longer foraging trips, which together explain the large drop observed in chick production between banded and non-banded birds during our study decade (80 versus 47 chicks produced; Supplementary Fig. 1). Moreover, decreasing breeding success in seabirds increases their dispersal²⁵, and dispersal of penguins is still studied almost exclusively using flipper-banded birds. The question then arises about the significance of such data, as dispersal may not be representative of actual population dispersal in free-living penguins and may therefore constitute a serious bias.

Finally, banded and non-banded penguins were differently affected by climate. Environmental conditions (Southern Oscillation index (SOI) and sea surface temperature (SST)) are known to affect penguins through changes in food availability (abundance or distribution), compelling individuals to forage for longer periods to reach sufficient body condition when conditions are warmer²⁶. In this regard, banded penguins arrived later to breed than did non-banded birds and even more so in years of lower SOI (that is, warmer phases of El Niño/Southern Oscillation²⁷; Supplementary Table 2 and Supplementary Fig. 4). Additionally, when compared with that of non-banded penguins, the breeding success of banded birds was similar in years of late arrivals (difference of -0.03 , $P = 0.52$, $N = 53$ birds), lower (albeit not significantly) in years of early arrivals (difference of -0.13 , $P = 0.32$, $N = 29$) and much lower in years of intermediate arrivals (difference of -0.19 , $P = 0.01$, $N = 70$). Thus, over a single year, differences may not be apparent. Food availability at sea may be so poor in a given year that even non-handicapped birds fail in large numbers. For instance, in 2007 (a year of late arrivals) both banded and non-banded king penguins similarly failed to breed. In contrast, in years of very favourable environmental conditions, the environmental pressure on banded birds may be so weak that they may compensate for the extra cost inflicted by banding, which would explain the absence of (or weakness in) difference observed between banded and non-banded birds in favourable years such as 2004, 2005 and 2006. This accords with data on banded Adélie penguins, whose increase in foraging duration varied according to the year⁸, and on African penguins, which seem to be negatively affected by banding only during periods of reduced prey availability²⁸. Similarly, we found that the population growth rates of banded and non-banded king penguins did not respond in the same way to variations in SST. Indeed, the relation between SST and the difference in population growth rates between the two groups was well approximated by a quadratic relationship, highlighting that this difference was most apparent in years of intermediate SST values (Fig. 2).

Thus, our decade-long monitoring demonstrates the detrimental effect of flipper banding and its interaction with climate on the major life-history traits of king penguins (Fig. 3). The effects of extensive banding schemes on penguin populations can on ethical grounds no longer be neglected, and studies considering population trajectories with regard to climate change seriously need to reconsider the biases inherent in studies using flipper-banded birds.

METHODS SUMMARY

King penguins from a breeding colony on Possession Island ($46^{\circ} 25' S$, $51^{\circ} 45' E$; Crozet archipelago) were monitored from 1998 to 2008 using automatic

identification (see ref. 26 for details). One hundred birds were randomly sampled in their breeding area and 50 of them were fitted with a metal flipper band⁷ (banding every second bird).

We established breeding propensity and phenology by analysing bird movements between their breeding area and the sea²⁹. Breeding propensity and success were defined as binary variables respectively depending on whether or not a bird attempted to breed and whether or not a breeding bird fledged its chick. In king penguins, egg laying is not synchronized within the colony, owing to a breeding cycle lasting over a year^{21,29}. Here we present results of early breeders only, because they are the ones that most affect population reproductive success. We investigated survival as a function of decline in animal presence. Population growth rates were estimated as dominant eigenvalues of transition matrices³⁰ and simulated according to varying SST.

Monthly fluctuations in SOI, SST and chlorophyll *a* concentration ([Chl *a*]) both around Crozet archipelago and around the polar front were the three environmental descriptors used in this study²⁶. SOI was used as a global index of El Niño/Southern Oscillation, whereas SST and [Chl *a*] were used as proxies of the prey availability of important feeding locations of king penguins.

We computed all statistics using the R statistical environment (version R-2.8.0). Survival was investigated using a Cox proportional hazards model with right censoring. Changes in cumulative survival over time were determined using breakpoints in segmented regression analysis. We analysed breeding data using mixed models computed with the individual as a random effect, enabling us to account for repeated measures, as birds were tracked over multiple breeding seasons. Years were added as random effects when we did not investigate differences between years.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 August; accepted 29 October 2010.

- Richardson, A. J. & Poloczanska, E. S. Ocean science: under-resourced, under threat. *Science* **320**, 1294–1295 (2008).
- Tynan, C. T. Ecological importance of the southern boundary of the Antarctic Circumpolar current. *Nature* **392**, 708–710 (1998).
- Croxall, J. P., Trathan, P. N. & Murphy, E. J. Environmental change and Antarctic seabird populations. *Science* **297**, 1510–1514 (2002).
- Durant, J. M. *et al.* The pros and cons of using seabirds as ecological indicators. *Clim. Res.* **39**, 115–129 (2009).
- Jackson, S. & Wilson, R. P. The potential costs of flipper-bands to penguins. *Funct. Ecol.* **16**, 141–148 (2002).
- Fallow, P., Chiaradia, A., Ropert-Coudert, Y., Kato, A. & Reina, R. Flipper-bands modify the short-term diving behaviour of Little penguins. *J. Wildl. Mgmt* **73**, 1348–1354 (2009).
- Gauthier-Clerc, M. *et al.* Long-term effects of flipper-bands on penguins. *Proc. R. Soc. Lond. B* **271**, S423–S426 (2004).
- Dugger, K., Ballard, G., Ainley, D. & Barton, K. Effects of flipper-bands on foraging behavior and survival of Adélie penguins (*Pygoscelis Adeliae*). *Auk* **123**, 858–869 (2006).
- Boersma, D. P. & Rebstock, G. A. Effects of double bands on Magellanic penguins. *J. Field Ornithol.* **81**, 195–205 (2010).
- Boersma, D. P. & Rebstock, G. A. Flipper-bands do not affect foraging-trip duration of Magellanic penguins. *J. Field Ornithol.* **80**, 408–418 (2009).
- Hindell, M. A., Lea, M. A. & Hull, C. L. The effects of flipper bands on adult survival rate and reproduction in the royal penguin, *Eudyptes schlegeli*. *Ibis* **138**, 557–560 (1996).
- Barbraud, C. & Weimerskirch, H. Emperor penguins and climate change. *Nature* **411**, 183–186 (2001).
- Jenouvrier, S. *et al.* Demographic models and IPCC climate projections predict the decline of an emperor penguin population. *Proc. Natl Acad. Sci. USA* **106**, 1844–1847 (2009).
- Croxall, J. P., McCann, T. S., Prince, P. A. & Rothery, P. in *Antarctic Ocean and Resources Variability* (ed. Sahrhage, D.) 261–285 (Springer, 1988).
- Le Maho, Y. *et al.* Undisturbed breeding penguins as indicators of changes in marine Resources. *Mar. Ecol. Prog. Ser.* **95**, 1–6 (1993).

16. May, R. M. Ethics and amphibians. *Nature* **431**, 403 (2004).
17. Lebreton, J. D. & Clobert, J. in *Bird Population Studies: their Relevance to Conservation and Management* (eds Perrins, C. M., Lebreton, J. D. & Hiron, G. J. M.) 105–125 (Oxford Univ. Press, 1991).
18. Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, 1992).
19. Pfister, C. A. Patterns of variance in stage-structured populations: evolutionary predictions and ecological implications. *Proc. Natl Acad. Sci. USA* **95**, 213–218 (1998).
20. Gaillard, J.-M., Festa-Bianchet, M. & Yoccoz, N. G. Population dynamics of large herbivores: variable recruitment with constant adult survival. *Trends Ecol. Evol.* **13**, 58–63 (1998).
21. Weimerskirch, H., Stahl, J. C. & Jouventin, P. The breeding biology and population dynamics of king penguins *Aptenodytes patagonica* on the Crozet Islands. *Ibis* **134**, 107–117 (1992).
22. Olsson, O. Seasonal effects of timing and reproduction in the king penguin: a unique breeding cycle. *J. Avian Biol.* **27**, 7–14 (1996).
23. Koslov, A. N. & Tarverdiyeva, I. Feeding of different species of Myctophidae in different parts of the Southern Ocean. *J. Ichthyol.* **29**, 160–167 (1989).
24. Groscolas, R. & Robin, J.-P. Long-term fasting and re-feeding in penguins. *Comp. Biochem. Physiol. A* **128**, 645–655 (2001).
25. Boulonier, T., McCoy, K. D., Yoccoz, N. G., Gasparini, J. & Tveraa, T. Public information affects breeding dispersal in a colonial bird: kittiwakes cue on neighbours. *Biol. Lett.* **4**, 538–540 (2008).
26. Le Bohec, C. *et al.* King penguin population threatened by Southern Ocean warming. *Proc. Natl Acad. Sci. USA* **105**, 2493–2497 (2008).
27. Deser, C. & Wallace, J. M. El Niño events and their relation to the southern oscillation: 1925–1986. *Geophys. Res. Lett.* **92**, 14189–14196 (1987).
28. Wolfaardt, A. C. & Nel, D. C. in *The Rehabilitation of Oiled African Penguins: A Conservation Success Story* (eds Nel, D. C. & Whittington, P. A.) 18–24 (Birdlife South Africa, 2003).
29. Descamps, S., Gauthier-Clerc, M., Gendner, J. P. & Le Maho, Y. The annual breeding cycle of unbanded king penguins *Aptenodytes patagonicus* on Possession Island (Crozet). *Avian Sci.* **2**, 87–98 (2002).
30. Caswell, H. *Matrix Population Models: Construction, Analysis and Interpretation* 72–76, 300–306 (Sinauer, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Trathan for his comments on the paper. We are grateful to C. Salmon for his help in pre-analysing data. We thank M. Ballesteros, C. Bricaud, N. Chatelain, G. Conan, S. Descamps, J. Dutel, C. Gilbert, S. Gravier, A. Hergott, G. Kuntz, N. Lambert, T. Lebard, N. Lecomte, J. Legrand, S. Mangin, V. Mosch, S. Quéméneur, A. Simon, E. Taquet and C. Villemin, for their help with field work. We also thank I. Durant for her help with chlorophyll data. This work was supported by the Institut Polaire Français–Paul-Emile Victor, the Fondation Bettencourt-Schueller, the Fondation des Treilles and the Centre National de la Recherche Scientifique (Programme Zone Atelier de Recherches sur l'Environnement Antarctique et Subantarctique), and grants from the European Commission (Marie Curie, to C.L.B.), the Norwegian Research Council through the Match/Mismatch and Ecosystem project (to J.M.D.) and the YGGDRASIL programme (to C.S.).

Author Contributions C.S. did the analyses and co-wrote the paper; C.L.B. helped in the analyses, the organization of the paper and the writing; Y.L.M. designed the study and co-wrote the paper; V.A.V. provided ideas on the analyses and co-wrote the paper; N.G.Y. proposed one of the analyses and helped with statistics; J.M.D. supplied ideas on the analyses and the organization of the paper; M.G.-C. and N.C.S. added some very useful comments and modifications to the manuscript; D.B. ran some pre-analyses; and Y.-H.P. provided climatic data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.S. (claire.saroux@c-strasbourg.fr) or Y.L.M. (yvon.lemaho@iphc.cnrs.fr).

METHODS

We confirm that all work followed approval by an ethics committee and conforms to the legal requirements of the country in which it was carried out, including those relating to conservation and welfare.

Demographic survey. Our study was conducted on king penguins breeding in the colony of La Grande Manchotière at Possession Island (46° 25' S, 51° 45' E) on Crozet archipelago. We used automatic identification and data-logging systems (the ANTAVIA system³¹) to follow, from 1998 to 2008, 100 king penguins implanted with a passive transponder tag under the skin of their leg. The transponder tags weigh 0.8 g and have no known adverse effects. They were shown not to affect survival of king penguins³² or breeding success, recruitment or survival of tits³³. Furthermore, concerns about infections should be minimal, as transponder tags were kept sealed sterile in iodine capsules (Betadine) and were removed from the capsules only by the process of injecting them into the bird. Moreover, Vétédine soap and alcoholic antiseptic solutions were used to disinfect the skin and the injecting needle before each insertion. Flesh wounds did not seem infected thereafter (personal observations on recaptured birds). Penguins were randomly sampled in their breeding area (only breeding birds were taken, to make sure that they were mature birds). Fifty of them were also fitted with a metal flipper band⁷, with the simple rule of banding every second bird to randomize the treatment. Automatic identification systems allow for continuous, automatic data collection of bird movements into and out of the colony. The system is completed by video recording of the passageways over the antennas. Banded birds were thus monitored during the whole study, and only one bird lost its flipper band. This bird was not considered in further analyses.

Biological variables. The breeding propensities and phenologies of the birds were established by analysing recorded movements of the 100 studied penguins between the breeding area and the sea²⁹. We assumed that those birds which were recorded leaving the colony went to sea. The specificity of the king penguins' breeding cycle allows us to determine whether and when birds attempted to breed, and whether and when they failed (stop in the sex-specific shifts). We were thus able to obtain arrival and laying dates, lengths of sojourns at sea and breeding variables, over all years and for each bird.

The sex of the birds was established by analysing the chronology of sex-specific incubating shifts²⁹. Because sex had no significant effect on the date of arrival at the colony or on the propensity to breed, gender was not incorporated in our models. The presence of couples in the sample was controlled to avoid double-counting the same reproductive event and to keep independent our data on each individual. The data recorded during 1998 (that is, the year the birds were marked) were not included, to avoid eventual bias induced by handling.

Breeding success, sea trips and survival. We defined breeding propensity as a binary variable considering the onset of a breeding cycle (breeding propensity equals 1 if the bird attempted reproduction and equals 0 otherwise, that is, if no egg was laid). We defined breeding success as successful chick fledging for birds that laid an egg (breeding success equals 1 if the chick is fledged and equals 0 otherwise).

Lengths of sea trips were investigated for all birds still incubating or brooding, whether successfully fledging their chick or not. Different shifts have different mean durations²⁹ and we thus separated them into different groups: one group per incubation shift, one for the first guard trip, one for all subsequent guard trips pooled together and a last one for all post-guard trips. To be able to run a single model on foraging trips, we standardized trips per group and then pooled them all together.

Yearly chick production was investigated as the number of fledglings (that is, 14-month-old chicks) produced per individual present in the colony each year. It combines breeding success and propensity into a single parameter representative of yearly reproduction at the population level. The total number of chicks produced over the 10-year period, integrating both breeding and survival over the number of penguins present in the colony, was also studied to give a better representation of the success of the considered sample (that is, banded or non-banded).

Survival was investigated as a function of decline in bird presence at the colony. We conducted analyses on both annual and overall survival. Overall survival corresponds to the number of studied birds present at the colony at the end of the period divided by the number of studied birds present at the colony at the beginning of the period.

Early breeders and late breeders. Because the king penguins' breeding cycle lasts for more than 1 year (around 14 months on Crozet archipelago^{21,29,34}), bird arrival at the colony depends on the success and timing of the previous year's breeding attempt. The laying period of king penguins therefore extends for over four months, with two peak periods^{34,35}: one for 'early breeders' (before 1 January) and another for 'late breeders' (after 1 January). To account for the dependence on the previous breeding attempt, we separated our data into two timing groups

and conducted separate analyses on them. The first corresponded to penguins that did not fledge a chick the previous breeding season (failed or skipped reproduction) and which were thus free to arrive early in the summer (early group), and the second corresponded to birds that succeeded in fledging a chick the previous breeding season and which were consequently late breeders the subsequent season (late group).

As birds in the late group have a very small impact on the chick production of the colony, we did not present their data in our breeding analyses. Their late arrival almost always results in breeding failure as they do not breed early enough to allow their chick to meet normal phases of the chick growth period^{21,22}. Moreover, the pool of these birds is very small in comparison with the early group, and the production of chicks and, therefore, the renewal of the population is thus mostly dependent on the success of early breeders.

Environmental descriptors. The three environmental descriptors used in this study were the SOI, calculated from the monthly fluctuation in the air pressure difference between Tahiti and Darwin, Australia; the SST (in degrees Celsius); and the concentration of chlorophyll ([Chl *a*], in milligrams per cubic metre) (see ref. 26 for details). Negative SOI values indicate a warm phase of El Niño/Southern Oscillation²⁷. SST and [Chl *a*] were used as a proxy of prey availability both around the breeding site and near the polar front, which are two locations known to affect the breeding of king penguins in the Crozet archipelago.

Growth rate estimates and simulations. To integrate breeding success and survival into one single parameter, we established population growth rates for both banded and non-banded groups, as the dominant eigenvalues of stage-structured population transition matrices³⁰. Different population matrix structures were tested and the obtained growth rates were almost identical, seeming insensitive to this structure. For simplicity, we present results of only one model. Briefly, we used a four-stage structured matrix with three first stages of immature birds and a last one of breeding adults. This supposes that every penguin began breeding at age five (average breeding age of the colony, unpublished data). An example of such a matrix is

$$M = \begin{pmatrix} 0 & 0 & 0 & \text{GBS} \\ \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \text{Adult survival} \end{pmatrix}$$

where GBS stands for global breeding success (that is, the product of breeding proportion and breeding success) and α represents annual juvenile survival.

As birds were followed only from adult age in this study, we fixed annual juvenile survival for both populations at 0.89 (C. Saraux *et al.*, unpublished observations, where the authors studied the return rates and survival of juveniles in king penguins). Breeding success is affected by the SST of the area around Crozet archipelago as far south as the polar front and survival is affected by SST at the MIZ with a 2-year lag²⁶. We computed two different models explaining breeding success in terms of SST around Crozet, one for banded birds and one for non-banded birds. Similarly, survival was modelled using the SST at the MIZ for banded and non-banded birds. Significant relationships were obtained in each of these four cases and were implemented in the matrix of each group as follows:

$$M_{\text{non-banded}} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{1 + e^{4.1207\text{SST}_{\text{Cro}} - 30.6312}} \\ \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \frac{1}{1 + e^{1.845\text{SST}_{\text{MIZ}} - 8.6676}} \end{pmatrix}$$

$$M_{\text{banded}} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{1 + e^{3.386\text{SST}_{\text{Cro}} - 24.981}} \\ \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \frac{1}{1 + e^{1.1304\text{SST}_{\text{MIZ}} - 5.5565}} \end{pmatrix}$$

SSTs averaged over the two different areas were highly correlated ($\rho = 0.97$, $P < 0.001$), and we thus decided to let both SSTs vary the same way, differing only by a constant: $\text{SST}_{\text{Cro}} = \text{SST}_{\text{MIZ}} + \text{mean}(\text{SST}_{\text{Cro}} - \text{SST}_{\text{MIZ}})$. Finally, we simulated the variation of these population growth rates in relation to varying SST. Standard errors of growth rates were calculated using Caswell's method³⁰, by adding standard errors of all parameters of the matrix, these having previously been calculated bootstrapping over 1,000 repetitions of the models used in the matrix.

Statistics. All statistics were computed using the R-2.8.0 statistical environment (R Development Core Team, 2008). Survival was investigated using a Cox proportional hazards model with right censoring. Changes in survival over time were determined

using breakpoints from segmented regression analysis ('segmented' package). Breeding data was analysed using a mixed-effect model fitted using maximum likelihood ('lme4' package³⁶). Generalized linear mixed models were computed with the individual (bird) as a random effect, enabling us to account for repeated measures, as birds were tracked over multiple breeding seasons. The most appropriate model was selected by using the Akaike information criterion. Variables were considered significant for $P < 0.05$.

31. Gendner, J. P., Gauthier-Clerc, M., Le Bohec, C., Descamps, S. & Le Maho, Y. A new application for transponders in studying penguins. *J. Field Ornithol.* **76**, 138–142 (2005).
32. Froget, G., Gauthier-Clerc, M., Le Maho, Y. & Handrich, Y. Is penguin banding harmless? *Polar Biol.* **20**, 409–413 (1998).
33. Nicolaus, M., Bouwman, K. & Dingemanse, N. Effect of PIT tags on the survival and recruitment of Great tits. *Ardea* **96**, 286–292 (2009).
34. Barrat, A. Quelques aspects de la biologie et de l'écologie du Manchot Royal (*Aptenodytes patagonicus*) des îles Crozet. *Com. Natl Fr. Rech. Antarct.* **40**, 9–51 (1976).
35. Stonehouse, B. The king penguin *Aptenodytes patagonica* of South Georgia. I. Breeding behaviour and development. *Falkl. Isl. Depend. Surv. Sci. Rep.* **23**, 1–81 (1960).
36. Bates, D. & Maechler, M. lme4: Linear mixed-effects models using Eigen and Eigen. *CRAN - Package lme4* (<http://CRAN.R-project.org/package=lme4>) (2009).

Distributed biological computation with multicellular engineered networks

Sergi Regot^{1*}, Javier Macia^{2*}, Núria Conde^{1,2}, Kentaro Furukawa³, Jimmy Kjellén³, Tom Peeters¹, Stefan Hohmann³, Eulàlia de Nadal¹, Francesc Posas¹ & Ricard Solé^{2,4,5}

Ongoing efforts within synthetic and systems biology have been directed towards the building of artificial computational devices¹ using engineered biological units as basic building blocks^{2,3}. Such efforts, inspired in the standard design of electronic circuits^{4–7}, are limited by the difficulties arising from wiring the basic computational units (logic gates) through the appropriate connections, each one to be implemented by a different molecule. Here, we show that there is a logically different form of implementing complex Boolean logic computations that reduces wiring constraints thanks to a redundant distribution of the desired output among engineered cells. A practical implementation is presented using a library of engineered yeast cells, which can be combined in multiple ways. Each construct defines a logic function and combining cells and their connections allow building more complex synthetic devices. As a proof of principle, we have implemented many logic functions by using just a few engineered cells. Of note, small modifications and combination of those cells allowed for implementing more complex circuits such as a multiplexer or a 1-bit adder with carry, showing the great potential for re-utilization of small parts of the circuit. Our results support the approach of using cellular consortia as an efficient way of engineering complex tasks not easily solvable using single-cell implementations.

Engineered living cells have been designed to perform a broad variety of functions^{8–16} but, with few exceptions, complex computational constructs (such as comparators, bit adders or multiplexers) are difficult to obtain and reuse¹⁷. Moreover, cell-cell communication requirements rapidly grow with circuit complexity, thus limiting the combinatorial potential of the constructs. One way of overcoming these difficulties is to use cellular consortia¹⁸ based on the idea that external communication between cells in populations involving either single^{19–21} or multiple^{22–24} cell types would perform functions difficult to be implemented using individual strains. Here, we apply this view to a novel distributed approach based on a reusable, sparse design of synthetic circuits.

A small library of engineered cell types with restricted connections among them was generated, each cell responding to one/two inputs (Fig. 1a–c). The basic two-input and one-output engineered functions include the AND and the inverted IMPLIES (N-IMPLIES, Fig. 1d), which allow implementing any Boolean function. Moreover, some cells define one-input, one-output function (Fig. 1e). The output of each cell type is either a diffusible wiring molecule or the desired output. In contrast with previous works using synthetic consortia^{18–22}, we have not used cell-cell feedbacks. Instead, cells only respond to an external input and to a single diffusible molecule acting as a wire.

The computation is determined by: (1) the number of cells C involved, (2) the specific function implemented by each engineered cell and (3) the location of cells within the network (see Supplementary Information and Supplementary Fig. 1 for details). Crucially, we allow different engineered cells to produce the output signal, which is thus distributed.

Moreover, each cell can be modulated by external inputs, which can either trigger the production of a signal or its inhibition.

The combinatorial nature of our approach is highlighted by calculating, for each C , the number of functions that can be implemented¹⁹. We have analysed all possible functions with two and three inputs versus C with our approach (see Supplementary Information and Supplementary Fig. 2) and found that most can be constructed using $C = 2–5$ different cells. For instance, in response to three inputs, just three cells results in more than 100 functions and exceed 200 using four cell types. The number of extracellular wires using this approach is significantly lower compared to other standard approximations (Supplementary Fig. 3 and Supplementary Information). With three inputs, over 100 different logical functions can be achieved with only two wires and almost all are obtained with just three to four wires (Supplementary Fig. 3).

As a proof of principle that distributed computation can be implemented *in vivo*, we created a library of engineered yeast cells. Each cell

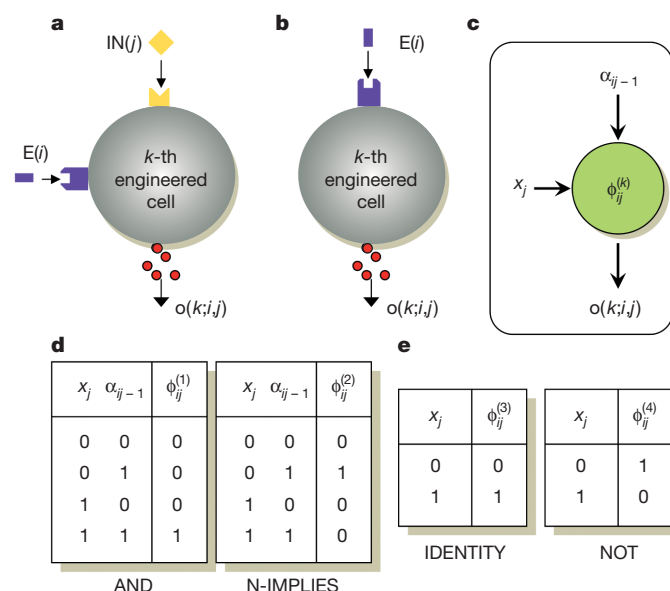


Figure 1 | Basic engineered cells (cell types). **a, b**, Cells can receive signals from other cells (IN) and external sources (E) (**a**) or just from external sources (**b**). Cells can also produce diffusible output molecules. **c**, Representation of the cell behaviour is summarized, where each cell c_{ij} responds to two different inputs; external input (x_j) and a signalling molecule (wire) from another cell (α_{ij-1}). The response of k -th cell type $\phi_{ij}^{(k)}$ can be the production of a new wiring molecule (α_{ij}) or the final output. The k -th cells respond to the presence of signals through some Boolean function $\Phi_{ij}^{(k)}$ with $k = \{1, 2, 3, 4\}$ defining the resulting Boolean output. **d, e**, The four basic functions implemented in our study are displayed.

¹Cell signaling unit, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra (UPF), E-08003 Barcelona, Spain. ²ICREA-Complex Systems Laboratory, Universitat Pompeu Fabra (UPF), E-08003 Barcelona, Spain. ³Department of Cell and Molecular Biology/Microbiology, University of Gothenburg, Box 462, 40530 Gothenburg, Sweden. ⁴Santa Fe Institute, Santa Fe, New Mexico 87501, USA. ⁵Institut de Biologia Evolutiva, CSIC-UPF, Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain.

*These authors contributed equally to this work.

responds to an extracellular stimulus (for example, NaCl, doxycycline, galactose, oestradiol) and/or the presence of a wiring molecule (for example, yeast pheromone). The output of the cells was monitored as the expression of a reporter construct under the control of the *FUS1* promoter (for example, green fluorescent protein, GFP). See Supplementary Fig. 4 for relevant genotype and the logic function of each cell of the library is. The ability of cells to respond to external stimuli (inputs) was monitored by fluorescence in single cell (fluorescence-activated cell sorting, FACS) and normalized to the maximal number of cells able to produce output signal (see Fig. 2a and Supplementary Information). Each cell type has been characterized by its ability to respond to the corresponding stimuli (Fig. 2b and Supplementary Fig. 5).

We then implemented all standard 2-input logic functions by combining just a few engineered cell types. We initially designed a basic circuit with an AND logic (Fig. 3a) involving two cell types responding to two stimuli (NaCl and oestradiol) and using a pheromone (alpha factor) as a wiring molecule. The presence of NaCl stimulates Cell 1 to produce pheromone (IDENTITY) that is received by Cell 2. In addition, Cell 2 has the ability to sense another external input (oestradiol) and it is competent, via the production and activation of the Fus3 mitogen-activated protein kinase (MAPK), to produce the final output. Only in the presence of the two inputs the final outcome was produced (Fig. 3a). Similarly, a NOR gate was implemented using a

different pair of cell types in which each cell responded to a particular stimulus (doxycycline and 6a, an inhibitor of Fus3as kinase) with yeast pheromone as a wiring. Only in the absence of both stimuli there was positive output (Fig. 3b).

Next, we designed two completely different circuits that involved the use of three independent engineered cell types, by reusing cells from our previous AND and NOR circuits. The first three-cell circuit was an OR logic gate in which the two inputs are NaCl and galactose. In this circuit, engineered Cell 1 and 5 are IDENTITY functions. They respond to the presence of NaCl (input 1) or galactose (input 2) to produce the wiring molecule that induces output production in Cell 6 (GFP). The presence of any input (galactose or NaCl) generated a positive output as it corresponds to an OR gate (Fig. 3c). Similarly, a NAND gate was designed using doxycycline and glucose as inputs. Cell 3 and Cell 5 display NOT logic. Both secreted pheromone in the absence of stimuli. Cell 6 responded to the presence of pheromone from either Cell 3 or Cell 5 inducing a fluorescent output. As expected, only the presence of both stimuli generates the output (Fig. 3d). This illustrates how to increase computational complexity at low cost. Other circuits can be easily built through reuse (Supplementary Fig. 6). Of note, the N-IMPLIES circuit can be implemented in a single cell (Supplementary Fig. 6a) or by combining cells with different logics (Supplementary Fig. 6b). Using other consortia, we obtained the AND, NOR, OR, NAND, XNOR and XOR gates. However, they can be

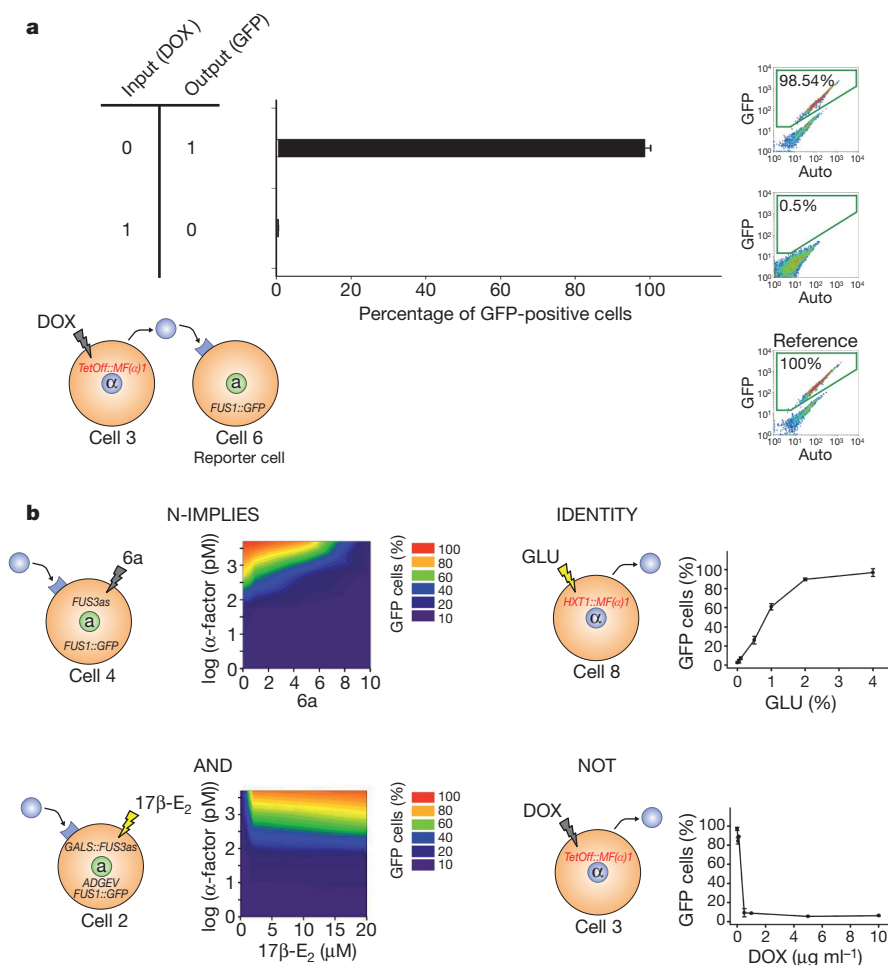


Figure 2 | *In vivo* analyses of engineered cells. **a**, Quantification of single cell computational output. Truth table and schematic representation of a cell with a NOT logic (see Supplementary Information for complete genotype). The NOT function is implemented in Cell 3, and the reporter cell (Cell 6) is used to quantify alpha factor production *in vivo*. Doxycycline (DOX) was added as indicated and cells were analysed by FACS. Data are expressed as the

percentage of GFP-positive cells versus cells treated with pheromone. Results represent the mean \pm s.d. of three independent experiments. **b**, Transfer functions of basic logic cells. Schematic representation of cells implementing N-IMPLIES, AND, IDENTITY and NOT functions. Indicated cells were treated with indicated input concentrations (2 inputs, left; 1 input, right). 17β-E₂, oestradiol; GLU, glucose.

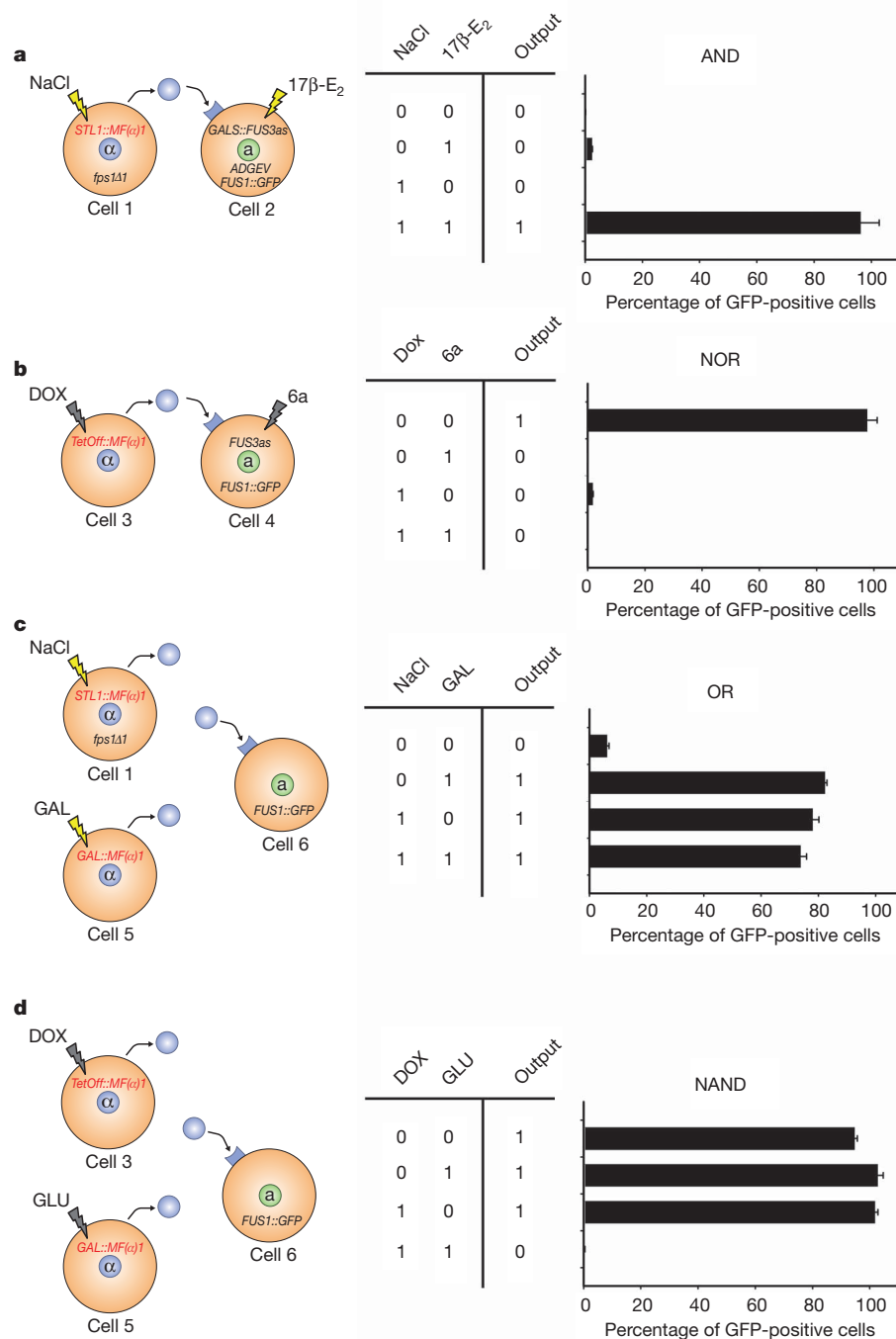


Figure 3 | Engineered cells to implement different logic gates *in vivo*.

a, Truth table and schematic representation of cells in the AND circuit (see Supplementary Information for complete genotype). Cells were mixed proportionally and inputs (NaCl and oestradiol) were added at the same time. **b**, Panel ordered as in **(a)** following NOR logic. Indicated cells were treated

using as inputs doxycycline and 6a. **c**, OR gate. Indicated cells were treated using as inputs 0.4 M NaCl and 2% galactose (GAL). **d**, NAND gate. Indicated strains were treated using as inputs doxycycline and 2% glucose. Data represent the mean and standard deviation of three independent experiments.

implemented using the same inputs for all of them (that is, doxycycline and glucose) (Supplementary Fig. 7a–f). This supports that our approach is adaptable and that multiple functions can be constructed from a small library of reusable cells.

We then analysed the long-term dynamical response of our AND circuit under changing inputs (Supplementary Fig. 7a). We have found that once the circuit is turned on, it can maintain maximal signal for periods beyond 9 h in the presence of stimuli (Supplementary Fig. 8a). Furthermore, once the system has been established, it responds equally well (at least eight generations) while the culture is maintained to a log phase (Supplementary Fig. 8b). We have also experimentally addressed

network responsiveness to dynamic changes by means of a microfluidic device containing the cellular types of the AND circuit and have exposed it to changes in the input signals over time. The system was able to dynamically respond to re-stimulation after GFP inactivation (Supplementary Fig. 8c).

Our system can be selectively switched off and partially reprogrammed. For instance, the inhibition of the intracellular signal transduction in Cell 2 of the AND gate, blocks the positive outcome of the circuit (Supplementary Fig. 9a). More interestingly, when reprogramming is applied to complex multicellular circuits, different computations can be obtained with little effort. For instance, when a reprogramming

molecule (glucose) is added to the OR gate shown in Supplementary Fig. 9b it works as IDENTITY for NaCl (input 2). In this context, despite the fact that multicellular circuits might seem more complex, their easy reuse and combination actually makes them more appropriate in many situations.

Finally, we were able to engineer complex circuits by re-using our previous designs. One of them is the multiplexer MUX2to1 circuit that selects one of different input signals and forwards the selected input into a single output. This circuit, if designed in a single cell would be difficult to implement *in vivo* (see Supplementary Fig. 10a). However, using distributed computation, the circuit can be assembled from just three engineered cell types responding to three input signals and a single wiring molecule (Supplementary Fig. 10b). In addition, we also implemented a MUX2to1 circuit that contains four cell types but uses two independent wiring molecules (α -factor from *Saccharomyces cerevisiae* and the α -factor from *Candida albicans*). Cell 10 and Cell 13 respond to doxycycline and produce each one of the wiring

molecules. Cell 12 responds to oestradiol and *S. cerevisiae* pheromone whereas Cell 15 responds to galactose and *C. albicans* pheromone. The final output (GFP) is generated by Cell 12 and Cell 15. Here, although the complexity of the circuit required a differential output to eight different input combinations, the *in vivo* results clearly showed that the computation of the three inputs yielded the expected response (Fig. 4a). A second complex circuit, the 1-bit adder with carry, was built by combining XOR and AND gates that respond to the same input (doxycycline and glucose) with two wiring molecules (α -factor from *S. cerevisiae* and *C. albicans*). In addition, output cells express different reporter proteins, a green reporter (adder) or a red reporter (carry) (*FUS1::GFP* or *FUS1::mCherry* respectively) allowing to detect the outcome of the carry and adder in the same culture. The system responds as an XOR gate (green columns) but presence of the two stimuli induces led to a 1-bit carrier (red columns) (Fig. 4b).

Possible applications as well as some caveats will need to be addressed in future work (such as scalability^{25,26}, strategies for reducing

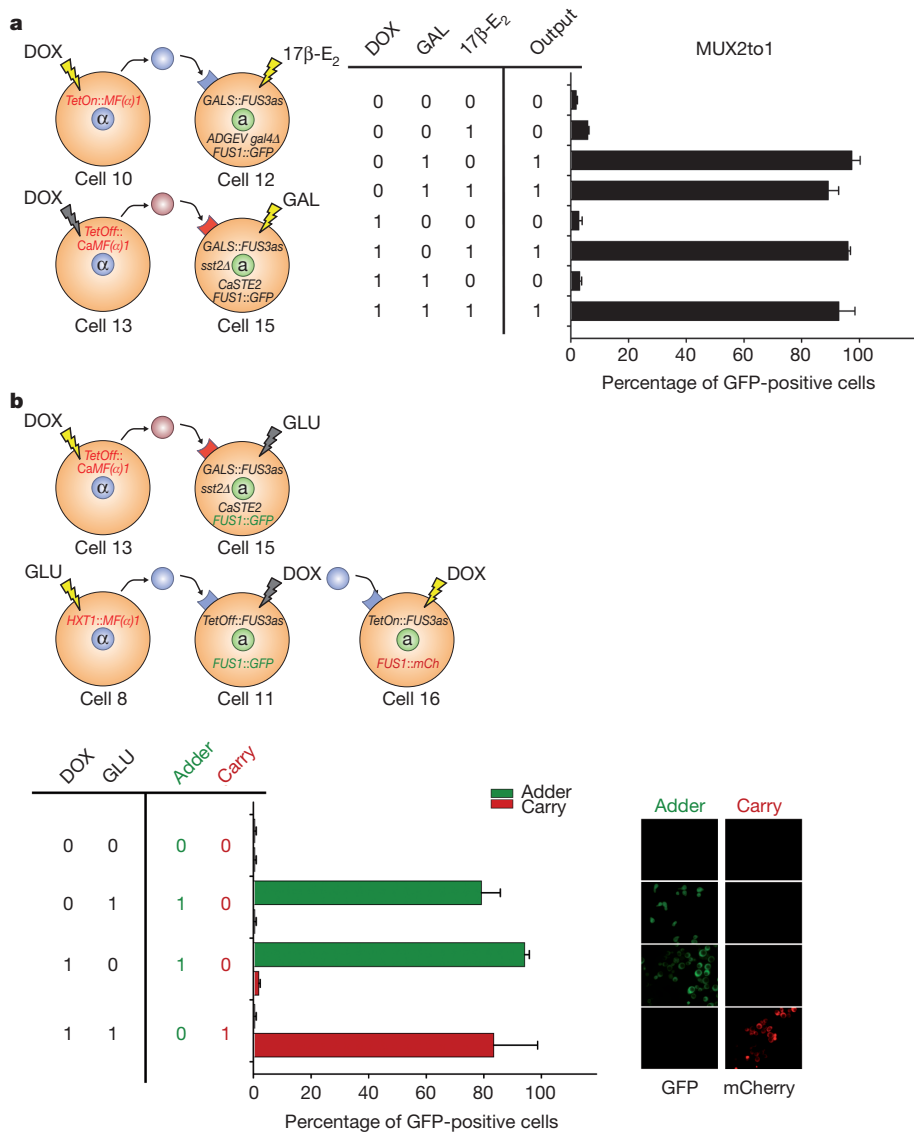


Figure 4 | Design and *in vivo* implementation of a multiplexer (MUX2to1) and 1-bit adder with carry. **a**, Truth table and schematic representation of the cells used in the MUX2to1. Indicated cells were treated using doxycycline (selector) and the inputs oestradiol and/or 2% galactose. Data are expressed as the percentage of GFP-positive cells using a sample treated with either *S. cerevisiae* or *C. albicans* alpha factor as a reference for Cell 12 or Cell 15, respectively. **b**, Truth table and schematic representation of cells used for 1-bit

adder with carry. Four cells with two wiring systems that respond to glucose and doxycycline with an XOR logic were combined with an extra cell that respond to same stimuli but with an AND logic in which instead of GFP, mCherry was expressed as output. The final outcome was measured as in Fig. 3a. Green bars indicate the adder output (GFP) whereas red bars represent the carry bit (mCherry). GFP and mCherry images of cells are shown (right panels). Data represent the mean and standard deviation of three independent experiments.

potential crosstalk³ and robustness to noise²⁷). However, the results reported here show that distributed computation using consortia is a powerful strategy to build complex synthetic constructs, opening a new door to reusable, reprogrammable complex circuits (Supplementary Fig. 11).

METHODS SUMMARY

Complete list of engineered yeast strains and plasmids is described in Supplementary Information. Computational output detection was done in single cells by flow cytometry (FACScalibur, Becton Dickinson) and the dynamical inputs responses to different circuits were analysed in a microscopy based microfluidic platform. See Supplementary Information for details about experimental methods.

Received 18 February; accepted 18 November 2010.

Published online 8 December 2010.

1. Bashor, C. J. & Horwitz, A. A. Peisajovich, S. G., & Lim, W. A. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living. *Ann. Rev. Biophys.* **39**, 515–537 (2010).
2. Amos, M. *Cellular Computing* (Oxford University Press, 2004).
3. Andrianantoandro, E., Basu, S., Karig, D. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, E1–E14 (2006).
4. Bray, D. Protein molecules as computational elements in living cells. *Nature* **376**, 307–312 (1995).
5. Hjeltnelt, A., Weinberger, E. D. & Ross, J. Chemical implementation of neural networks and Turing machines. *Proc. Natl Acad. Sci. USA* **88**, 10983–10987 (1991).
6. McAdams, H. H. & Shapiro, L. Circuit simulation of genetic circuits. *Science* **269**, 650–656 (1995).
7. Arkin, A. & Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **67**, 560–578 (1994).
8. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
9. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
10. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
11. Voigt, C. A. Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17**, 548–557 (2006).
12. Kramer, B. P., Fischer, C. & Fussenegger, M. BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol. Bioeng.* **87**, 478–484 (2004).
13. Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
14. Greber, D. & Fussenegger, M. An engineered mammalian band-pass network. *Nucleic Acids Res.* **38**, e174 (2010).
15. Weber, W. et al. A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proc. Natl Acad. Sci. USA* **105**, 9994–9998 (2008).
16. Rinaudo, K. et al. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnol.* **25**, 795–801 (2007).
17. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
18. Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26**, 483–489 (2008).
19. Brenner, K., Karig, D. K., Weiss, R. & Arnold, F. H. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl Acad. Sci. USA* **104**, 17300–17304 (2007).
20. Danino, T., Mondragón-Palmino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
21. You, L., Cox, R. S. III, Weiss, R. & Arnold, F. H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
22. Shou, W., Ram, S. & Vilar, J. M. G. Synthetic cooperation in engineered yeast populations. *Proc. Natl Acad. Sci. USA* **104**, 1877–1882 (2007).
23. Weber, W., Daoud-El Baba, M. & Fussenegger, M. Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proc. Natl Acad. Sci. USA* **104**, 10435–10440 (2007).
24. Balagaddé, F. K. et al. A synthetic *Escherichia coli* predator–prey ecosystem. *Mol. Syst. Biol.* **4**, 187 (2008).
25. Lucks, J. B., Qi, L., Whitaker, W. R. & Arkin, A. P. Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol.* **11**, 567–573 (2008).
26. Chin, J. W. Programming and engineering biological networks. *Curr. Opin. Struct. Biol.* **16**, 551–556 (2006).
27. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Subirana and S. Ovejas for technical support and S. Pellet and M. Peter for their help in setting up the microfluidics platform, and K. Kuchler for the *FUS1*-mCherry construct. S.R. is recipient of a FPU fellowship (Spanish Government). This work was supported by grants from the James McDonnell Foundation to R.S., the MICINN (BIO2009-07762 and FIS2009-12365); Consolider Ingenio 2010 programme (grant CSD2007-0015), from the ESF (ERAS-CT-2003-980409) FP6 as part of a EURYI scheme award (www.esf.org/euryi) to F.P. and the CELLCOMPUT (FP6) project to F.P., R.S. and S.H., and FP7 UNICELLSYS grant (#201142) to F.P. and S.H., and The Santa Fe Institute to R.S.; F.P. and R.S. laboratories are also supported by the Fundación Marcelino Botín (FMB). F.P. is recipient of an ICREA Acadèmia (Generalitat de Catalunya).

Author Contributions All authors shared all the phases of the work. J.M. and R.S. developed the theoretical background for multicellular computing. Circuits were designed by S.R., J.M., N.C., E.N., F.P. and R.S.; S.R., N.C., K.F., J.K. and T.P. did the experimental designs. J.M., F.P., E.N., S.H. and R.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.S. (ricard.sole@upf.edu) or F.P. (francesc.posas@upf.edu).

Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'

Alvin Tamsir¹, Jeffrey J. Tabor² & Christopher A. Voigt²

Computation underlies the organization of cells into higher-order structures, for example during development or the spatial association of bacteria in a biofilm^{1–3}. Each cell performs a simple computational operation, but when combined with cell–cell communication, intricate patterns emerge. Here we study this process by combining a simple genetic circuit with quorum sensing to produce more complex computations in space. We construct a simple NOR logic gate in *Escherichia coli* by arranging two tandem promoters that function as inputs to drive the transcription of a repressor. The repressor inactivates a promoter that serves as the output. Individual colonies of *E. coli* carry the same NOR gate, but the inputs and outputs are wired to different orthogonal quorum-sensing 'sender' and 'receiver' devices^{4,5}. The quorum molecules form the wires between gates. By arranging the colonies in different spatial configurations, all possible two-input gates are produced, including the difficult XOR and EQUALS functions. The response is strong and robust, with 5- to >300-fold changes between the 'on' and 'off' states. This work helps elucidate the design rules by which simple logic can be harnessed to produce diverse and complex calculations by rewiring communication between cells.

Boolean logic gates integrate multiple digital inputs into a digital output. Electronic integrated circuits consist of many layered gates. In cells, regulatory networks encode logic operations that integrate environmental and cellular signals^{6–8}. Synthetic genetic logic gates have been constructed, including those that perform AND, OR and NOT functions^{9–12}, and have been used in pharmaceutical and biotechnological applications^{13,14}. Multiple gates can be layered to build more complex programs^{15–17}, but it remains difficult to predict how a combination of circuits will behave on the basis of the functions of the individuals^{11,18}. Here we have compartmentalized a simple logic gate into separate *E. coli* strains and use quorum signalling to allow communication between the strains⁵. Compartmentalizing the circuit produces more reliable computation by population-averaging the response. In addition, a program can be built from a smaller number of orthogonal parts (for example transcription factors) by re-using them in multiple cells.

NOR and NAND gates are unique because they are functionally complete. That is, any computational operation can be implemented by layering either of these gates alone¹⁹. Of these, the NOR gate is the simplest to implement using existing genetic parts. A NOR gate is 'on' only when both inputs are 'off' (Fig. 1a). We designed a simple NOR gate by adding a second input promoter to a NOT gate²⁰. Tandem promoters with the same orientation drive the expression of a transcriptional repressor (Fig. 1b). Tandem promoters are common in prokaryotic genomes²¹. This is expected to produce an OR function; however, interference between the promoters can occur (Supplementary Figure 3). The repressor turns off a downstream promoter, which serves as the output of the gate. Both the inputs and the output of this gate are promoters; thus, multiple gates could be layered to produce more complex operations.

Each logic gate is encoded in separate strains of *E. coli*. Acyl homoserine lactone (AHL) cell–cell communication devices are used as

signal-carrying 'wires' to connect the logic gates encoded in different strains^{4,5,22}. Gates are connected in series where the output of the first gate is the expression of the AHL synthase (*Pseudomonas aeruginosa* PAO1 LasI or RhlI). AHL diffuses through the cell membrane and binds to its cognate transcription factor (*P. aeruginosa* PAO1 LasR or

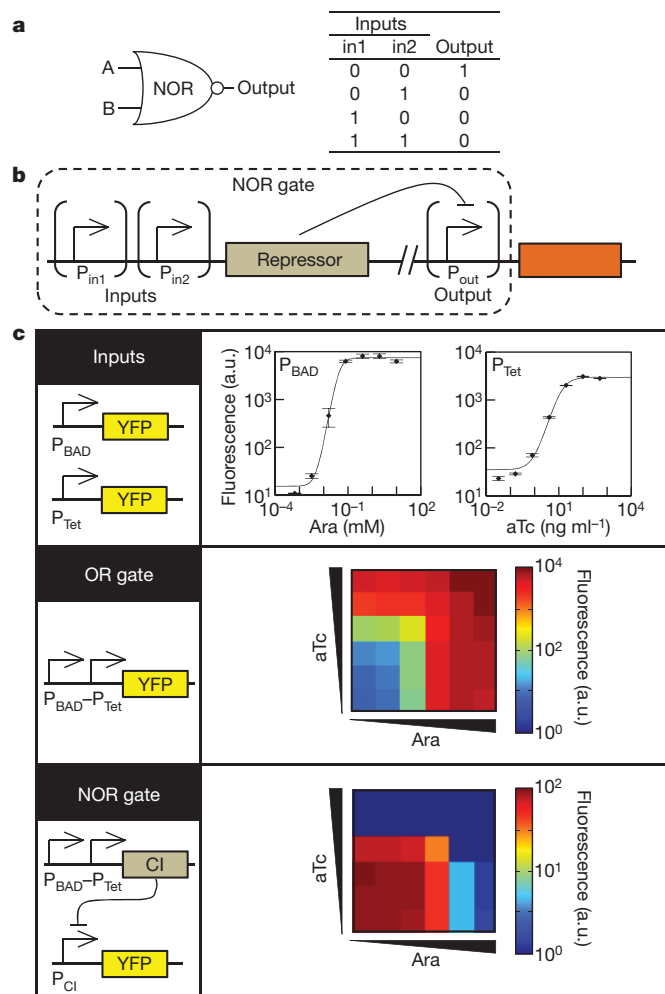


Figure 1 | The genetic NOR gate. **a**, Symbol, truth table (**a**) and genetic diagram (**b**) of the NOR gate. **c**, The transfer function is defined as the output as a function of input at steady state. The transfer functions of P_{BAD} and P_{Tet} (top), the $P_{BAD}-P_{Tet}$ tandem promoter (middle), and the NOR gate (bottom) are shown. The inducer concentrations for the tandem promoter and NOR gate characterizations are 0, 0.0005, 0.005, 0.05, 0.5 and 5 mM Ara (squares from left to right) and 0, 0.025, 0.25, 2.5, 25 and 250 ng ml⁻¹ aTc (squares from bottom to top). Fluorescence values and their error bars are calculated as mean \pm s.d. from three experiments. a.u., arbitrary units.

¹Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. ²Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94158, USA.

RhlR). The promoter that is turned on by the transcription factor is used as the input to the next logic gate. These systems have been used previously to program cell–cell communication and have been shown to have little cross-talk⁴. Analogous to a series of electrical gates arrayed on a circuit board, compartmentalization of genetic gates in individual cells allows them to be added, removed or replaced simply by changing the spatial arrangement of the *E. coli* strains.

The stepwise construction of a NOR gate with P_{BAD} and P_{Tet} as the input promoters and yellow fluorescent protein (YFP) as the output gene is shown in Fig. 1c. P_{BAD} and P_{Tet} are activated in the presence of arabinose (Ara) and anhydrotetracycline (aTc), respectively. The individual transfer functions of P_{BAD} and P_{Tet} are measured using flow cytometry (Fig. 1c). An OR gate is constructed by placing the P_{BAD} and P_{Tet} promoters in tandem. P_{BAD} – P_{Tet} demonstrates OR logic with 7,000-fold induction between the ‘off’ state (–Ara, –aTc) and the ‘on’ state (+Ara, +aTc). Finally, to convert the OR gate into a NOR gate, the CI-repressor gene is placed under the control of P_{BAD} – P_{Tet} and YFP is expressed from a second plasmid under the control of the CI-repressible P_R promoter. Whereas the OR gates have some characteristics of fuzzy logic, the NOR gates are nearly digital (Fig. 1c).

These OR and NOR gates use promoters as inputs. This feature imparts modularity to the gates; in other words, they can be engineered to respond to different inputs by replacing the promoters. To investigate this, we swapped the input promoters of the logic gates. Figure 2 shows the characterization data for three different tandem promoters: P_{BAD} – P_{Tet} , P_{BAD} – P_{Las} and P_{Tet} – P_{Las} . The promoter P_{Las} is activated by the quorum signal 3OC12-HSL²² (*N*-3-oxo-dodecanoyl-homoserine lactone). These gates perform as the additive combination of the individual transfer functions of the two input promoters and the CI-repressor NOT gate. The predicted transfer function for the six logic gates shown in Fig. 2 matched the experimental results. One tandem promoter, P_{Tet} – P_{BAD} , did not function as predicted (Supplementary

Figure 3). The failure observed for P_{Tet} – P_{BAD} probably arises from some position-dependent interference. This could be the result of the effects of DNA looping, the occlusion of transcription-factor-binding sites or changes in the ratio or stability of output messenger RNAs, among other effects.

Complex logic can be designed using layers of simpler gates. An XOR gate is built with three NOR gates and a buffer gate (Fig. 3a). The output of an XOR gate is ‘on’ only when either (but not both) inputs are ‘on’. Four strains, each carrying a different logic gate, are used to construct an XOR circuit. The strains are spotted onto an agar plate in the spatial arrangement required to perform this function (Fig. 3b). Cell 1 carries a NOR gate that uses Ara and aTc as inputs and expresses LasI as the output. This allows cell 1 to be wired to the NOR gates in cells 2 and 3 by means of 3OC12-HSL. Cells 2 and 3 use Ara and aTc as their second inputs, respectively. Similarly, the output of the NOR gates in cells 2 and 3 is RhlI, which produces C4-HSL²² (*N*-butyryl-homoserine lactone). Cell 4 acts as a buffer gate and integrates the outputs from cells 2 and 3 by responding to C4-HSL. The output of a buffer gate is ‘on’ only when the input is ‘on’. The complete circuit consisting of all four strains behaves as a digital XOR gate with respect to the two inputs (Ara and aTc; Fig. 3c, d). Each intermediate colony performs its digital logical operations appropriately, as tested by replacing each output gene with YFP (Fig. 3c).

We constructed a small library of strains that act as simple logic gates, most of which are components of the XOR gate (Fig. 4a). Circuit diagrams showing how all of the sixteen possible two-input logic gates can be constructed using the library are shown in Fig. 4b. Each circuit diagram is reproduced by the spatial arrangement of the component strains. None of these circuits required additional genetic manipulation. The range of induction varies from 5-fold (XOR) to 335-fold (B gate). The dominant contribution to the dynamic range of the complete circuit is due to the intrinsic range of the final circuit (Supplementary Figure 7). For example, the XOR and NAND gates are limited by the output of P_{Rhl} . The addition of a NOT gate to this promoter increases the dynamic ranges of the EQUAL, AND, A IMPLY B and B IMPLY A gates, which is an effect described previously²³. No degradation in the signal is observed as a function of the number of layers.

The calculations are robust with respect to the distance between colonies and the time and density at which they are spotted (Supplementary Figure 10). This robustness is partially due to the population averaging that occurs, which reduces the effect of cell–cell variation. Despite the variability in the circuit response within a colony, this variability is effectively averaged and thus is not propagated to the next layer of the circuit. The use of chemical signals and population averaging could represent a common design rule for achieving computational operations robust enough to overcome the stochastic limitations of layered circuits in individual cells^{24,25}. Another source of robustness is the external clock that is implemented by delaying the spotting of colonies for each layer. Genetic computing is asynchronous and this may result in hazards, that is, transient incorrect outputs that occur as a result of mismatched delays in the circuit²⁶. This is apparent when circuits are measured in liquid culture, where the calculation is less robust with respect to timing and cell density (Supplementary Figure 12). To perform the calculation properly, all of the cells need to start in the ‘off’ state. As layered computation becomes more critical to the design of genetic programs, this will either require the implementation of a genetic clock²⁷ or the design of programs that are robust to asynchronous computation²⁸.

Cellular automata have been used to show how simple logic yields complex patterns in the organization of cells¹. These have been used to model biological pattern formation, development and complex collective behaviour^{3,29}. Here we demonstrate that a library of simple gates can be used to form more complex computational operations by linking the gates using diffusible chemical signals. The motif of multiple promoters in tandem driving the expression of a repressor is common

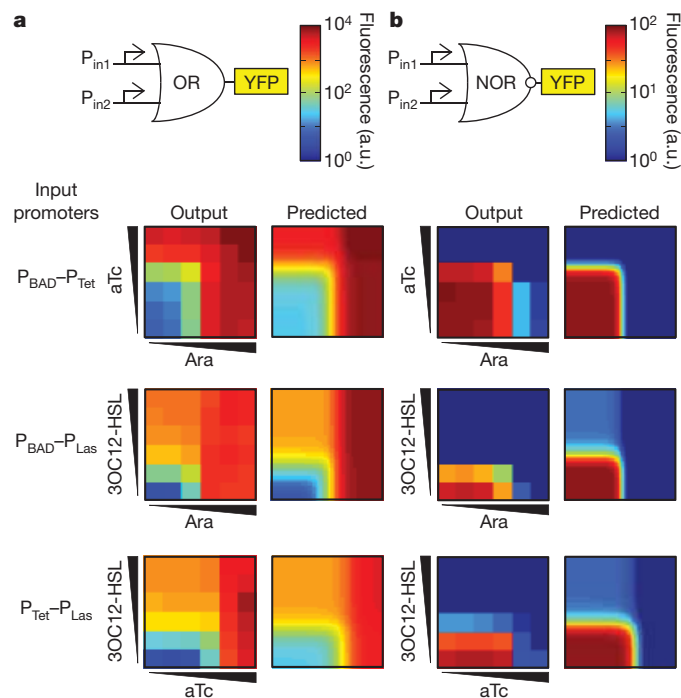


Figure 2 | Input modularity of the gates. **a**, Transfer functions for three OR gates (left) are compared with the predicted transfer function (right). The predicted transfer function is the simple sum of the transfer functions measured for the individual promoters (Supplementary Information). The Ara and aTc concentrations used are the same as in Fig. 1 and those for 3OC12-HSL are 0, 0.001, 0.01, 0.1, 1 and 10 μ M (squares from bottom to top). **b**, Transfer functions for three NOR gates (left) are compared with the predicted transfer functions (right). The data represent means calculated from three experiments.

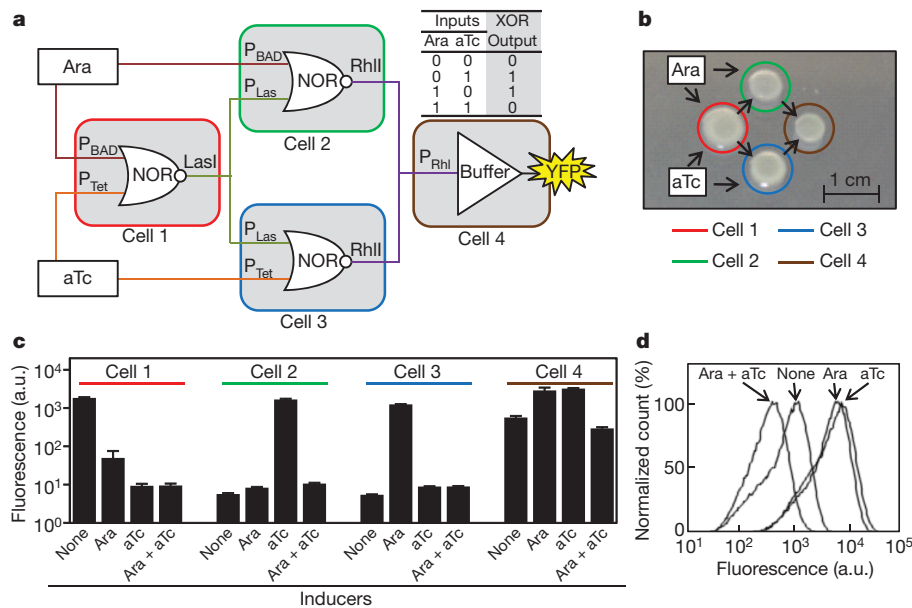


Figure 3 | Construction of an XOR gate by programming communication between colonies on a plate. **a**, Four colonies—each composed of a strain containing a single gate—are arranged such that the computation progresses from left to right, with the result of each layer communicated by means of quorum signals. The inputs (Ara and aTc) are added uniformly to the plate.

b, Spatial arrangement of the colonies. **c**, Each colony responds appropriately to the combinations of input signals. Fluorescence values and their error bars are calculated as mean \pm s.d. from three experiments. **d**, Cytometry data for the XOR gate (cell 4).

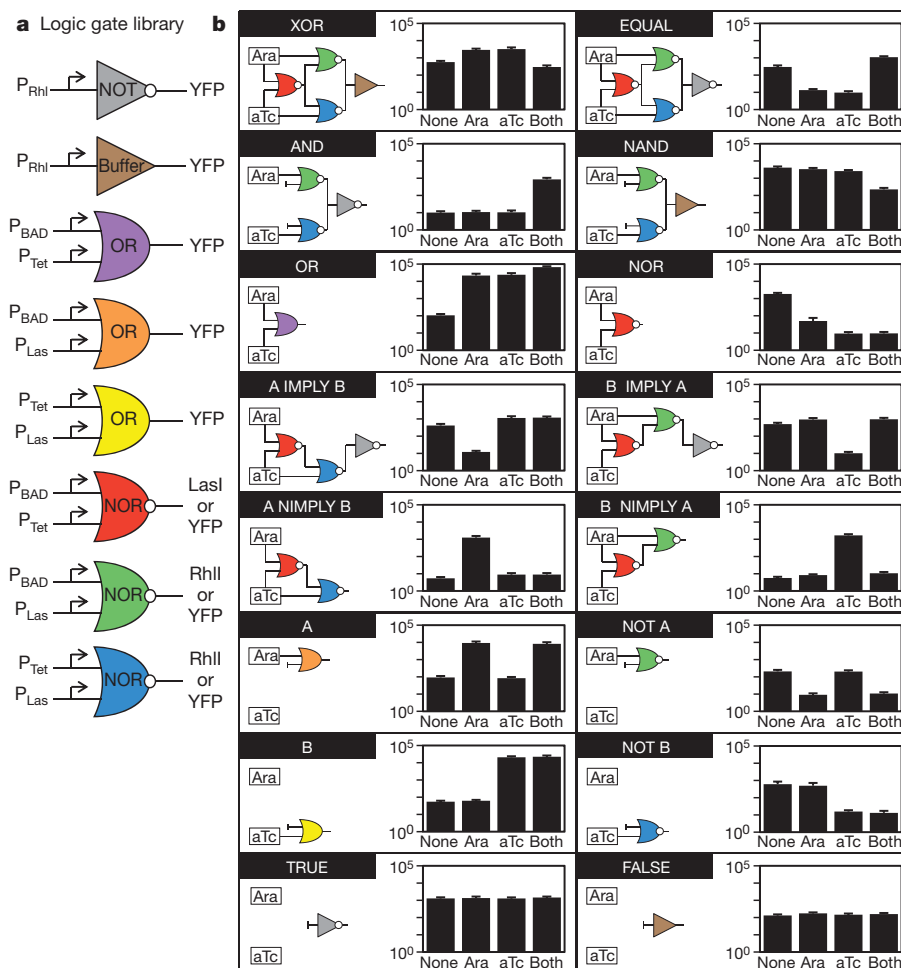


Figure 4 | Construction of all 16 two-input Boolean logic gates. **a**, Library of simple logic gates carried by different strains (corresponding to plasmids in Supplementary Table 5). **b**, Colonies containing different gates were spotted to mimic the spatial arrangement of each logic circuit (Fig. 3b). For each circuit,

the final colony was assayed by flow cytometry for all combinations of inducers added to the plate. The data correspond to the cytometry distributions in Supplementary Figure 6. Fluorescence values and their error bars are calculated as mean \pm s.d. from three experiments. NIMPLY, NOT IMPLY.

in genomes²¹, and the resulting NOR gates may represent a ubiquitous fundamental unit of biological computation. Although our current ability to create logic gates within a single cell is limited, it may ultimately be possible to encode more complex circuits in individual cells that are then linked by cell–cell communication, akin to logic blocks in field-programmable gate arrays³⁰. Together, these principles can be used in the engineering of biological systems to create increasingly complex functions.

METHODS SUMMARY

Strains, plasmids and media. All studies were performed using *E. coli* strain DH10B. Luria–Bertani (LB)–Miller medium (Difco 244610) was used for the assays. The antibiotics used were 50 µg ml⁻¹ chloramphenicol (Acros 227920250) and/or 50 µg ml⁻¹ kanamycin (Fisher BP906-5). The inducers used were arabinose (Sigma A3256), anhydrotetracycline (Fluka 37919) and 3OC12-HSL (Sigma O9139).

Transfer function characterization. Cells harbouring the appropriate plasmids were incubated in 3 ml of LB broth medium (37 °C, 250 r.p.m. shaking) in culture tubes without the presence of inducers for 18 h. The cultures were then diluted 200-fold into 200 µl fresh LB broth medium (supplemented with appropriate inducers) in a 96-well plate format and incubated for additional 14 h before finally being diluted 100-fold into PBS solution for cytometry analysis.

Plate assay of circuit function. The plate medium was prepared by pouring 12 ml of LB broth agar medium (1.5% agar (Difco 214030), 2.5% LB–Miller) supplemented with inducers (2 mM Ara and/or 500 ng ml⁻¹ aTc) into a 100-mm Petri dish (Fisher 08-757-13). Bacterial logic gates were ‘fabricated’ on the plate by spotting 1 µl overnight culture of appropriate bacterial strains (Supplementary Table 5) to mimic the spatial arrangement of each logic circuit. The distance between each two colonies was set at 7 mm in square grids. Spotting was done with 12-h delay from the previous layer’s spotting to ensure communication signals had propagated sufficiently. After 12 h from the last layer’s spotting, the whole output colony of the circuit was scraped using inoculating loops and diluted into 10 ml PBS solution for cytometry analysis.

Flow cytometry. All data contained at least 50,000 events, obtained using BD-FACS LSR2. Events were gated by forward and side scatter using MATLAB software. The geometric means of the fluorescence distributions were calculated. The autofluorescence value of *E. coli* DH10B cells harbouring no plasmid was subtracted from these values to give the fluorescence values reported in this study.

Received 2 March; accepted 11 October 2010.

Published online 8 December 2010.

1. Neumann, J. V. *The General and Logical Theory of Automata* (Wiley, 1951).
2. Turing, A. M. The chemical basis of morphogenesis. 1953. *Bull. Math. Biol.* **52**, 119–152 (discussion), 153–197 (1990).
3. Wolfram, S. *A New Kind of Science* 23–113 (Wolfram Media, 2002).
4. Brenner, K., Karig, D. K., Weiss, R. & Arnold, F. H. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl Acad. Sci. USA* **104**, 17300–17304 (2007).
5. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
6. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA* **101**, 4781–4786 (2004).
7. Niklas, K. J. The bio-logic and machinery of plant morphogenesis. *Am. J. Bot.* **90**, 515–525 (2003).

8. Morris, M. K., Saez-Rodriguez, J., Sorger, P. K. & Lauffenburger, D. A. Logic-based models for the analysis of cell signaling networks. *Biochemistry* **49**, 3216–3224 (2010).
9. Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A. & Alon, U. Plasticity of the cis-regulatory input function of a gene. *PLoS Biol.* **4**, e45 (2006).
10. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3**, 133 (2007).
11. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1466–1470 (2002).
12. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnol.* **25**, 795–801 (2007).
13. Weber, W. *et al.* A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proc. Natl Acad. Sci. USA* **105**, 9994–9998 (2008).
14. Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnol.* **27**, 465–471 (2009).
15. Lou, C. *et al.* Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol. Syst. Biol.* **6**, 350 (2010).
16. Tabor, J. J. *et al.* A synthetic genetic edge detection program. *Cell* **137**, 1272–1281 (2009).
17. Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
18. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nature Chem. Biol.* **5**, 842–848 (2009).
19. Scharle, T. W. Axiomatization of propositional calculus with Sheffer functors. *Notre Dame J. Formal Logic* **6**, 209–217 (1965).
20. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99**, 16587–16591 (2002).
21. Sneppen, K. *et al.* A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli*. *J. Mol. Biol.* **346**, 399–409 (2005).
22. Pesci, E. C., Pearson, J. P., Seed, P. C. & Igilewski, B. H. Regulation of las and rhl quorum sensing in *Pseudomonas aeruginosa*. *J. Bacteriol.* **179**, 3127–3132 (1997).
23. Karig, D. & Weiss, R. Signal-amplifying genetic circuit enables *in vivo* observation of weak promoter activation in the Rhl quorum sensing system. *Biotechnol. Bioeng.* **89**, 709–718 (2005).
24. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
25. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
26. Katz, R. H. & Borriello, G. *Contemporary Logic Design* 141–146 (Prentice Hall, 1994).
27. Danino, T., Mondragon-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
28. Clancy, K. & Voigt, C. A. Programming cells: towards an automated ‘genetic compiler’. *Curr. Opin. Biotechnol.* **21**, 572–581 (2010).
29. Ilachinski, A. *Cellular Automata: A Discrete Universe* 1–18 (World Scientific, 2001).
30. Raju, B. S. & Mullick, S. K. Programmable cellular arrays. *Int. J. Control* **14**, 1041–1061 (1971).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Mulyasmita and K. Temme for critical discussions. This work was supported by the National Science Foundation (SynBERC, NSF#0943385 and NSF Sandpit CCF-0943385) and the Office of Naval Research.

Author Contributions A.T. designed and performed the experiments, analysed the data and wrote the manuscript. J.J.T. designed experiments and edited the manuscript. C.A.V. designed experiments, analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.A.V. (cavoigt@picasso.ucsf.edu).

Integrative genomics identifies *LMO1* as a neuroblastoma oncogene

Kai Wang^{1*†}, Sharon J. Diskin^{2*}, Haitao Zhang^{1*}, Edward F. Attiyeh², Cynthia Winter², Cuiping Hou¹, Robert W. Schnepf², Maura Diamond², Kristopher Bosse², Patrick A. Mayes², Joseph Glessner¹, Cecilia Kim¹, Edward Frackelton¹, Maria Garriss², Qun Wang², Wendy Glaberson¹, Rosetta Chiavacci¹, Le Nguyen^{2,3,5}, Jayanti Jagannathan², Norihisa Saeki⁴, Hiroki Sasaki⁴, Struan F. A. Grant^{1,5,6}, Achille Iolascon^{7,11}, Yael P. Mosse^{2,5}, Kristina A. Cole^{2,5}, Hongzhe Li³, Marcella Devoto^{3,5,6,8}, Patrick W. McGrady⁹, Wendy B. London¹⁰, Mario Capasso^{7,11}, Nazneen Rahman¹², Hakon Hakonarson^{1,5,6} & John M. Maris^{2,5}

Neuroblastoma is a childhood cancer of the sympathetic nervous system that accounts for approximately 10% of all paediatric oncology deaths^{1,2}. To identify genetic risk factors for neuroblastoma, we performed a genome-wide association study (GWAS) on 2,251 patients and 6,097 control subjects of European ancestry from four case series. Here we report a significant association within LIM domain only 1 (*LMO1*) at 11p15.4 (rs110419, combined $P = 5.2 \times 10^{-16}$, odds ratio of risk allele = 1.34 (95% confidence interval 1.25–1.44)). The signal was enriched in the subset of patients with the most aggressive form of the disease. *LMO1* encodes a cysteine-rich transcriptional regulator, and its paralogues (*LMO2*, *LMO3* and *LMO4*) have each been previously implicated in cancer. In parallel, we analysed genome-wide DNA copy number alterations in 701 primary tumours. We found that the *LMO1* locus was aberrant in 12.4% through a duplication event, and that this event was associated with more advanced disease ($P < 0.0001$) and survival ($P = 0.041$). The germline single nucleotide polymorphism (SNP) risk alleles and somatic copy number gains were associated with increased *LMO1* expression in neuroblastoma cell lines and primary tumours, consistent with a gain-of-function role in tumorigenesis. Short hairpin RNA (shRNA)-mediated depletion of *LMO1* inhibited growth of neuroblastoma cells with high *LMO1* expression, whereas forced expression of *LMO1* in neuroblastoma cells with low *LMO1* expression enhanced proliferation. These data show that common polymorphisms at the *LMO1* locus are strongly associated with susceptibility to developing neuroblastoma, but also may influence the likelihood of further somatic alterations at this locus, leading to malignant progression.

Multiple somatically acquired chromosomal rearrangements, such as focal amplification of the *MYCN* oncogene or deletions at chromosome arms 1p or 11q, are each associated with an aggressive neuroblastoma phenotype². Although these somatically acquired genomic alterations are of clinical use as prognostic biomarkers, until recently little was known about the constitutional genetic events that initiate tumorigenesis. Highly penetrant gain-of-function mutations in the anaplastic lymphoma kinase (*ALK*) tyrosine kinase domain were recently identified as the major cause of familial neuroblastoma, and somatic mutations in this gene implicate it as a target for therapeutic intervention^{3–6}. In addition, a neuroblastoma GWAS identified common SNPs at 6p22 as being associated with susceptibility to aggressive neuroblastoma in sporadic cases⁷; follow-up association analysis on the clinically relevant group of patients with an aggressive tumour

phenotype indicated that common SNPs within *BARD1* also function as susceptibility variants⁸. Finally, our GWAS has also identified a common copy number variation at 1q21.1 being highly associated with neuroblastoma and probably playing a role in early tumorigenesis through disruption of a novel neuroblastoma breakpoint family gene (*NBPF23*)⁹. Taken together, it has become clear that the embryonal cancer neuroblastoma is genetically heterogeneous, and initiation of sporadically occurring disease requires multiple interacting genetic factors, including both sequence and copy number variants.

To identify additional genetic risk factors, we expanded our previous GWAS and analysed 1,627 neuroblastoma patients accrued through the North American-based Children's Oncology Group with 3,254 genetically matched control subjects of European ancestry (see Supplementary Methods). All subjects were genotyped using the Illumina HumanHap550 BeadChip with over 550,000 SNP markers; the genomic control inflation factor was 1.08 (Supplementary Fig. 1). Clusters of SNPs from three genomic loci reached genome-wide significance ($P < 5 \times 10^{-8}$; Fig. 1a), including two SNPs within *FLJ22536/FLJ44180* at the 6p22 locus (P values range from 2.46×10^{-14} to 3.25×10^{-13} ; Supplementary Table 1), nine SNPs within or nearby *BARD1* at the 2q35 locus (P values range from 3.05×10^{-13} to 9.69×10^{-9} ; Supplementary Table 2), each previously reported, and two SNPs within *LMO1* (LIM domain only 1), a newly identified neuroblastoma susceptibility locus at 11p15.4 (P values range from 5.12×10^{-10} to 2.83×10^{-8} ; Table 1 and Fig. 1b). Closer examination of the *LMO1* locus identified a total of four SNPs that show strong association signals ($P < 1 \times 10^{-4}$) with neuroblastoma (Table 1), which are in a moderate degree of linkage disequilibrium (Supplementary Fig. 2). We then examined each of the most significant SNPs from the 2q35, 6p22, 11p15.4 susceptibility loci and the 1q21.1 copy number variation. However, we did not find evidence for epistasis (Supplementary Tables 3 and 4), indicating that these susceptibility loci increase disease risk independently.

To replicate our findings, we examined the association results from an independent case series of 190 patients from the Children's Oncology Group and 1,507 control subjects, all of whom were genotyped on the Human610-Quad arrays. All four *LMO1* SNPs identified in the discovery effort showed the same direction of association in this replication cohort, with P values ranging from 1.01×10^{-5} to 0.058. To seek additional evidence of replication, we performed quantitative PCR-based genotyping of these four SNPs in a third independent case series

¹The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ²Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ³Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 19104, USA. ⁴Genetics Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan. ⁵Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 19104, USA. ⁶Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA. ⁷CEINGE Biotechnologie Avanzate, Naples 80145, Italy. ⁸Department of Experimental Medicine, University La Sapienza, Rome 00185, Italy. ⁹Department of Statistics, University of Florida and Children's Oncology Group, Gainesville, Florida, 32603 USA. ¹⁰Dana-Farber Children's Hospital Cancer Center and Children's Oncology Group, Boston, Massachusetts, 02115, USA. ¹¹Department of Biochemistry and Medical Biotechnology, University of Naples Federico II, Naples 80131 Italy. ¹²Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. [†]Present address: Zilkha Neurogenetic Institute, Department of Psychiatry and Preventive Medicine, University of Southern California, Los Angeles, California 90089, USA.

*These authors contributed equally to this work.

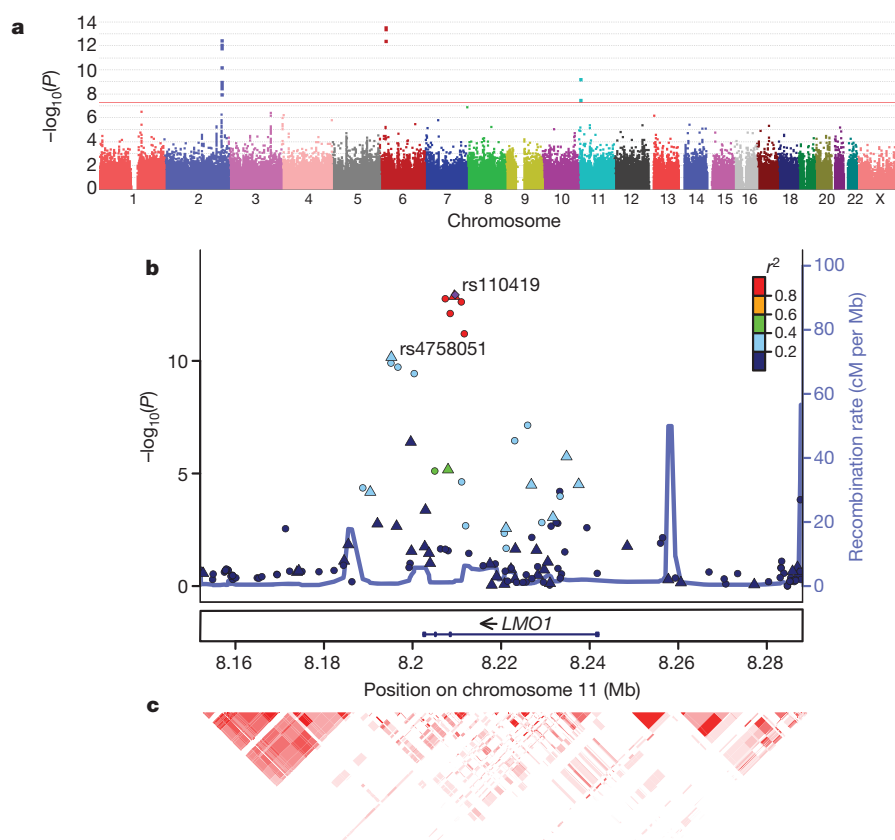


Figure 1 | Discovery of *LMO1* at 11p15.4 as a neuroblastoma susceptibility locus. **a**, Manhattan plot of GWAS results from the discovery cases series, with the red horizontal line representing genome-wide significance threshold ($P < 5 \times 10^{-8}$). **b**, Genomic position (National Center for Biotechnology Information build 36) of genotyped (triangles) and imputed (circles) SNPs. The

P values are calculated by combining discovery and replication case series with whole-genome genotypes, and SNPs are coloured based on their correlations with rs110419 (purple diamond). Estimated recombination rates from the HapMap data are overlaid. **c**, Degree of linkage disequilibrium between SNPs (as r^2 values) is represented by red colour intensity in the corresponding cells.

from UK, as well as the two most significant SNPs in a fourth independent case series from Italy. Combined analysis by the Cochran–Mantel–Haenszel method demonstrated that two of the four SNPs had *P* values that extend well beyond the genome-wide significance threshold (Table 1). Additionally, using the two cohorts with whole-genome genotype data (discovery cohort and US replication cohort), we performed genotype imputation at 11p15.4 and identified six additional genome-wide significant markers, the most significant being rs110420 ($P = 1.17 \times 10^{-13}$), which is in complete linkage disequilibrium ($r^2 = 1$ in HapMap CEU subjects (Utah residents with

ancestry from northern and western Europe) with the genotyped marker rs110419 (Fig. 1c and Supplementary Table 5).

We next determined if the *LMO1* genotypes were associated with a particular clinical phenotype and/or patient survival. Similar to the association pattern observed for the 6p22 and 2q35 (*BARD1*) loci^{7,8}, the risk alleles of *LMO1* were significantly associated with metastatic disease ($P = 0.0040$), advanced age (greater than 1 year, $P < 0.0001$) and a high-risk status by Children's Oncology Group criteria for treatment stratification² ($P = 0.0010$; Supplementary Tables 6 and 7). Consistent with this observation, the rs110419 risk allele was associated

Table 1 | Significantly associated SNPs at the *LMO1* locus on 11p15.4

| SNP† | Risk/non-risk allele | Discovery (HumanHap550)* | | | US replication (Human610)* | | | UK replication (TaqMan)* | | |
|------------|----------------------|---|--|-----------------------|---|--|----------------------|---|--|------------|
| | | Frequency of cases (<i>n</i> = 1,627) | Frequency of controls (<i>n</i> = 3,254) | <i>P</i> ‡ | Frequency of cases (<i>n</i> = 190) | Frequency of controls (<i>n</i> = 1,507) | <i>P</i> ‡ | Frequency of cases (<i>n</i> = 253) | Frequency of controls (<i>n</i> = 845) | <i>P</i> ‡ |
| rs4758051 | G/A | 0.51 | 0.45 | 2.8×10^{-8} | 0.55 | 0.45 | 2.1×10^{-4} | 0.51 | 0.45 | 0.039 |
| rs10840002 | A/G | 0.42 | 0.37 | 6.0×10^{-6} | 0.44 | 0.38 | 0.019 | 0.37 | 0.36 | 0.61 |
| rs110419 | A/G | 0.55 | 0.49 | 5.1×10^{-10} | 0.61 | 0.49 | 1.0×10^{-5} | 0.53 | 0.48 | 0.057 |
| rs204938 | C/T | 0.49 | 0.44 | 1.2×10^{-5} | 0.50 | 0.45 | 0.058 | 0.50 | 0.44 | 0.032 |
| | | | | | | | | | | |
| SNP† | Risk/non-risk allele | Italian replication (TaqMan)* | | | Combined | | | | | |
| | | Frequency of cases (<i>n</i> = 181) | Frequency of controls (<i>n</i> = 491) | <i>P</i> ‡ | CMH§ <i>P</i> | CMH OR (95% confidence interval)¶ | | | | |
| rs4758051 | G/A | 0.45 | 0.42 | 0.45 | 1.4×10^{-11} | 1.28 (1.19–1.37) | | | | |
| rs10840002 | A/G | — | — | — | 8.5×10^{-7} | 1.21 (1.12–1.30) | | | | |
| rs110419 | A/G | 0.49 | 0.41 | 0.004 | 5.2×10^{-16} | 1.34 (1.25–1.44) | | | | |
| rs204938 | C/T | — | — | — | 1.7×10^{-7} | 1.22 (1.13–1.31) | | | | |

*No deviations from Hardy–Weinberg equilibrium were observed ($P > 0.001$) in all cohorts.

†SNP: r^2 in controls between rs110419 and each of rs4758051, rs10840002 and rs204938 was 0.30, 0.17 and 0.29, respectively.

‡*P* values were calculated by allelic test.

§CMH, Cochran–Mantel–Haenszel test.

¶OR, odds ratio of risk allele.

with decreased event-free survival ($P = 0.0085$; Supplementary Table 8 and Supplementary Fig. 3) and overall survival ($P = 0.0217$; Supplementary Fig. 4). Taken together, these data suggest that common germline variants at *LMO1* are associated not only with predisposition to develop neuroblastoma, but also with a predilection to develop the more aggressive form of the disease. They emphasize that *LMO1* genetic variations are associated with a particular neuroblastoma phenotype; however, this does not indicate that these variants have prognostic significance for an individual with neuroblastoma.

The *LMO1* gene encodes a cysteine-rich transcriptional regulator with two LIM zinc-binding domains that is mainly expressed in the nervous system¹⁰. *LMO1* belongs to a protein superfamily encoded by four genes, including *LMO1*, *LMO2*, *LMO3* and *LMO4*. Multiple lines of evidence, including chromosomal translocation events and mouse models, strongly implicate this gene family in the aetiology of human cancer^{11–14}. Most provocatively, retroviral insertion of the corrective gene for X-linked severe combined immunodeficiency into the *LMO2* locus resulted in T-cell leukaemias in several participants in gene therapy trials¹⁵. *LMO4* represses the transcription of *BRCA1*, and dysregulation of *LMO4* expression has been implicated in the breast carcinogenesis^{16,17}. Finally, *LMO3* has been shown to act as an oncogene in neuroblastoma through the neuronal transcription factor *HEN2*¹⁸. We therefore postulated that the common variants at the 11p15.4 locus discovered here may increase disease risk through a *cis*-acting effect on the regulation of expression or function of *LMO1*, but we cannot exclude the potential for *trans*-acting influences on loci distant from the discovered common variants.

We next examined tumour DNA genotyped on the Illumina SNP arrays for 701 neuroblastomas using a detection algorithm for copy number designed for tumour samples¹⁹. We detected relative segmental gain (copy number changes at a given locus relative to whole-genome copy number changes) at *LMO1* in 87 out of 701 tumours (12.4%); this was particularly enriched in the high-risk group where the GWAS signal was most robust (Supplementary Fig. 5a). Most tumours with 11p gain showed a duplication of the entire chromosome p arm, but four tumours (approximately 5%) showed focal gain restricted to 11p15 including the *LMO1* locus (Supplementary Fig. 5b). These data demonstrate that *LMO1* is one of many genes showing somatic copy number gain on 11p, and here we used the GWAS data to prioritize it as a potential target of this somatically acquired chromosomal rearrangement.

We next examined whether somatic *LMO1* alterations were associated with neuroblastoma clinical phenotype and survival of patients (Supplementary Table 9). Gain of *LMO1* was significantly more common in tumours from patients with metastatic disease ($P < 0.0001$), advanced age (greater than 1 year, $P < 0.0001$), unfavourable pathological grade ($P = 0.0013$) and Children's Oncology Group high-risk classification ($P < 0.0001$). Gain of 11p was rarely observed in the *MYCN* amplified cases (Supplementary Table 9). Despite the strong association of 11p gain in cases without *MYCN* amplification, a known powerful adverse prognostic factor¹, *LMO1* gain was associated with decreased overall survival of patients ($P = 0.041$) (Supplementary Table 10 and Supplementary Figs 6 and 7).

To investigate how the neuroblastoma-associated *LMO1* alleles may contribute to tumour initiation and/or clinical phenotype, we next genotyped a set of human neuroblastoma-derived cell lines with Illumina SNP arrays, and measured messenger RNA (mRNA) and protein expression levels on the subset of lines without copy number changes at 11p to avoid the influence of somatic DNA alterations on gene expression. Cell lines with diploid 11p status and harbouring homozygous risk alleles showed significantly higher *LMO1* mRNA and protein expression than those with homozygous non-risk alleles (Fig. 2a and Supplementary Table 11). This trend held in an expanded set of 25 neuroblastoma cell lines with variable 11p status (Supplementary Fig. 8). To determine if this correlation existed in diagnostic tumour tissues, we next examined mRNA expression levels on a whole-genome Affymetrix expression microarray²⁰ in a subset of 61

neuroblastoma primary tumours from patients whose blood samples and primary tumours had both been genotyped on the Illumina SNP arrays. Among these 61 tumours, 13 harboured somatic gain of 11p. Considering both somatic and germline genotypes in the same linear regression model, we detected an association between *LMO1* copy number gains and increased *LMO1* expression ($P = 0.02$; Fig. 2b and Supplementary Table 12), as well as an association between rs110419 risk alleles and increased *LMO1* expression ($P = 0.022$; Fig. 2b). To refine the genotype-expression relationships further, we subsequently used quantitative PCR to measure *LMO1* expression in an additional set of 23 tumours without *LMO1* gain. We confirmed that the rs110419 risk allele is associated with *LMO1* expression ($P = 0.01$), independent of copy number changes (Fig. 2c). To determine whether a regulatory variant exists at a narrow promoter region of *LMO1*, we performed Sanger sequencing in 20 neuroblastoma cell lines but did not detect any potential causal variant (Supplementary Table 13). Examination of the 1000 Genomes Project data identified over 300 SNPs within or surrounding *LMO1* that are in moderate to strong linkage disequilibrium ($D' > 0.5$) with rs110419 (Supplementary Table 14); however, fine mapping of this region through resequencing will be required to identify whether any are causal *cis*-regulatory variants. Subsequent experimentation will be required to determine if causal DNA variations directly impact *LMO1* expression, and if somatic copy-number gain indeed is targeting *LMO1* for further increased expression in tumour cells.

As our germline and somatic genomic analyses implicated *LMO1* as a neuroblastoma oncogene, we next sought to determine the functional

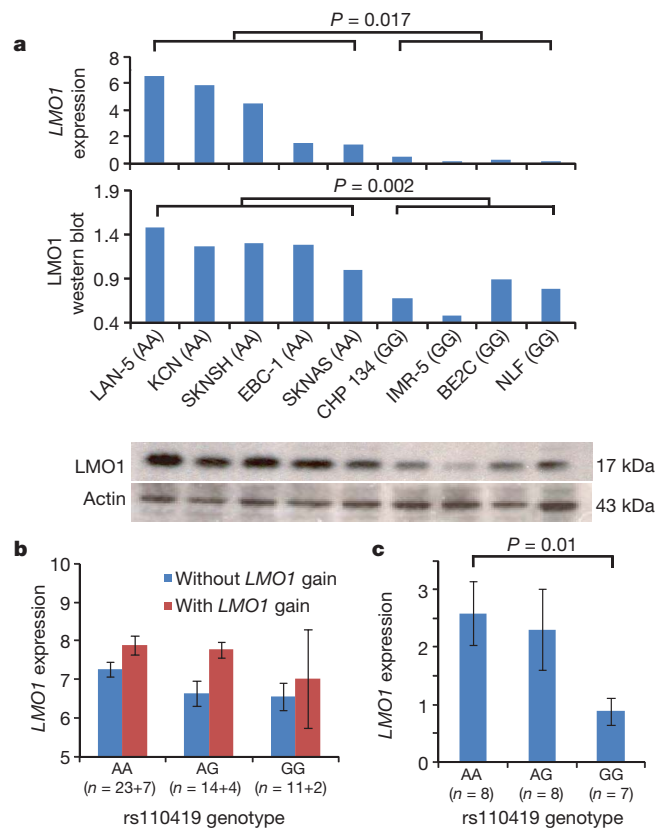


Figure 2 | *LMO1* germline genotypes and somatic copy number gains are associated with mRNA and protein expression. **a**, *LMO1* mRNA and protein expression in nine human neuroblastoma-derived cell lines are highly correlated with rs110419 genotype. **b**, Microarray-based expression profiling on 61 primary tumours confirms that *LMO1* gene expression is associated with both *LMO1* gain ($P = 0.02$, *t*-test) and risk genotypes ($P = 0.022$, linear regression). **c**, Quantitative PCR-based expression profiling of an independent set of primary neuroblastomas without *LMO1* gain confirms the same association. Error bars, s.e.m.

consequences of *LMO1* depletion or overexpression in a genotype- and expression-specific manner. First, after lentiviral-based shRNA infection of neuroblastoma cell lines, we were able to recover stable clones with 45–63% depletion of *LMO1* mRNA and protein (Fig. 3e). Cells with the homozygous neuroblastoma-associated genotype and high *LMO1* expression showed significantly decreased proliferation compared with mock-infected controls (Fig. 3a, b), whereas cells with homozygous non-risk alleles showed little phenotypic effect (Fig. 3c, d). Finally, to determine the cellular phenotypes of forced overexpression of *LMO1*, we stably overexpressed *LMO1* with approximately fourfold higher levels in the SK-N-BE2C cell line with low *de novo* *LMO1* expression, and detected significantly enhanced proliferation (Fig. 3f). Therefore it appears that inhibition of *LMO1* in cells expressing high levels of *LMO1* or activation of *LMO1* in cells with low levels of *LMO1* leads to pronounced phenotypes. Taken together, these data suggest that *LMO1* may function as an oncogene in a subset of human neuroblastomas.

In conclusion, here we have identified germline sequence variants at the *LMO1* locus that are robustly associated with neuroblastoma. We have applied an integrative genomics approach to demonstrate that common genetic polymorphisms associated with cancer predisposition may also mark regions of the genome prone to somatic alterations influencing tumour progression. Our data suggest that GWAS studies can identify previously undiscovered oncogenic drivers of a malignant

phenotype, especially when they occur in a region of the genome involved in large segmental rearrangements impacting hundreds of genes. In paediatric cancers such as neuroblastoma, the real translational potential of GWAS efforts may be in discovering therapeutic targets and predictive biomarkers of tumour aggressiveness.

METHODS SUMMARY

All genome-wide SNP genotyping for the discovery cohorts was performed using the Illumina HumanHap550 BeadChip at the Center for Applied Genomics at the Children's Hospital of Philadelphia. Multi-dimensional scaling was performed using PLINK version 1.06 on a subset of SNPs not in linkage disequilibrium to identify subjects of European ancestry, and all control subjects were genetically matched to patients. The first replication case series was genotyped by Illumina Human610 BeadChip, yet two additional replication case series were genotyped by TaqMan. Genotype imputation was performed by MACH (<http://www.sph.umich.edu/csg/abecasis/MACH/>) on discovery and replication case series with whole-genome genotypes. Alteration calls in tumour copy number were generated from data of SNP signal intensity by the OverUnder¹⁹. Survival analyses used the methods of Kaplan and Meier, with standard errors following the methods of Peto *et al.*²¹. For gene expression profiling by Affymetrix U95Av2 microarrays, the expression measures for each probe set was extracted and normalized using robust multi-array average protocols from raw CEL files. Association tests on genotype and expression were performed on log-transformed expression values by linear regression or *t*-test. For quantitative PCR on *LMO1*, TaqMan probes were purchased from Applied Biosystems with assay identity Hs00231133_m1. Relative expression of the target gene was determined by normalization to *HPRT1* using a standard curve method with ten serial dilutions according to the manufacturer's instructions. All quantitative PCR reactions were performed in triplicate with an ABI PrismTM 7900HT Sequence Detection System (Applied Biosystems). For the *LMO1* knockdown experiments, the lentiviral particles for shRNA knockdown were purchased from Santa Cruz, including copGFP Control Lentiviral Particles (catalogue number sc-108084) and *LMO1* shRNA(h) Lentiviral Particles (catalogue number sc-38025-v). Pooled clones of SK-N-BE2C cells with *LMO1* overexpression were created through stable transfection of full-length *LMO1* complementary DNA in pCDNA3.1 as previously described²².

Received 17 August 2009; accepted 22 October 2010.

Published online 1 December 2010.

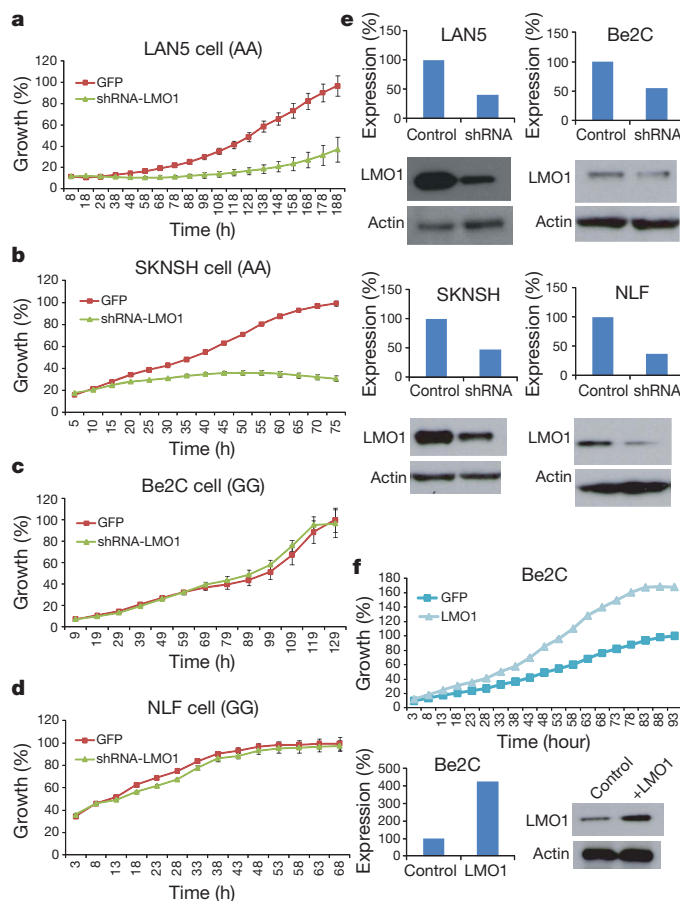


Figure 3 | Genetic manipulation of *LMO1* expression in neuroblastoma cell line models influences proliferative phenotype in an expression-specific manner. a–d, In cells with neuroblastoma risk alleles and higher *LMO1* expression levels, *LMO1* knockdown leads to inhibition of cellular proliferation. e, *LMO1* knockdown as measured by quantitative reverse-transcription PCR and western blot for experiments a–d. f, In SK-N-BE2C cells with non-risk alleles and low *LMO1* expression levels, forced overexpression of *LMO1* leads to enhanced cellular proliferation. Approximate fourfold overexpression of *LMO1* RNA and protein are shown. Error bars, s.e.m.

1. Maris, J. M. Recent advances in neuroblastoma. *N. Engl. J. Med.* **362**, 2202–2211 (2010).
2. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369**, 2106–2120 (2007).
3. Mosse, Y. P. *et al.* Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* **455**, 930–935 (2008).
4. Janoueix-Lerosey, I. *et al.* Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* **455**, 967–970 (2008).
5. Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971–974 (2008).
6. George, R. E. *et al.* Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature* **455**, 975–978 (2008).
7. Maris, J. M. *et al.* Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N. Engl. J. Med.* **358**, 2585–2593 (2008).
8. Capasso, M. *et al.* Common variations in *BARD1* influence susceptibility to high-risk neuroblastoma. *Nature Genet.* **41**, 718–723 (2009).
9. Diskin, S. J. *et al.* Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* **459**, 987–991 (2009).
10. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101**, 6062–6067 (2004).
11. Rabbitts, T. H. *LMO* T-cell translocation oncogenes typify genes activated by chromosomal translocations that alter transcription and developmental processes. *Genes Dev.* **12**, 2651–2657 (1998).
12. Rabbitts, T. H. *et al.* The effect of chromosomal translocations in acute leukemias: the *LMO2* paradigm in transcription and development. *Cancer Res.* **59**, 1794s–1798s (1999).
13. Fisch, P. *et al.* T-cell acute lymphoblastic lymphoma induced in transgenic mice by the *RBTN1* and *RBTN2* LIM-domain genes. *Oncogene* **7**, 2389–2397 (1992).
14. Neale, G. A., Reh, J. E. & Goorha, R. M. Disruption of T-cell differentiation precedes T-cell tumor formation in *LMO-2* (rhombotin-2) transgenic mice. *Leukemia* **11** (suppl. 3), 289–290 (1997).
15. Haein-Bey-Abina, S. *et al.* *LMO2*-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419 (2003).
16. Sum, E. Y. *et al.* The LIM domain protein *LMO4* interacts with the cofactor CtlP and the tumor suppressor *BRCA1* and inhibits *BRCA1* activity. *J. Biol. Chem.* **277**, 7849–7856 (2002).
17. Visvader, J. E. *et al.* The LIM domain gene *LMO4* inhibits differentiation of mammary epithelial cells *in vitro* and is overexpressed in breast cancer. *Proc. Natl Acad. Sci. USA* **98**, 14452–14457 (2001).

18. Aoyama, M. *et al.* LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. *Cancer Res.* **65**, 4587–4597 (2005).
19. Attiyeh, E. F. *et al.* Genomic copy number determination in cancer cells from single nucleotide polymorphism microarrays based on quantitative genotyping corrected for aneuploidy. *Genome Res.* **19**, 276–283 (2009).
20. Wang, Q. *et al.* Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. *Cancer Res.* **66**, 6050–6062 (2006).
21. Peto, R. *et al.* Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br. J. Cancer* **35**, 1–39 (1977).
22. Saeki, N. *et al.* GASDERMIN, suppressed frequently in gastric cancer, is a target of LMO1 in TGF-beta-dependent apoptotic signalling. *Oncogene* **26**, 6488–6498 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the Children's Oncology Group for providing most blood and tumour specimens and clinical and outcome data (U10-CA98543 and U10-CA98413) from neuroblastoma patients. We thank G. P. Tonini for providing neuroblastoma DNA samples in the Italian replication cohort. This work was supported in part by National Institutes of Health grant R01-CA124709 (to J.M.M.), the Giulio D'Angio Endowed Chair (J.M.M.), the Alex's Lemonade Stand Foundation (J.M.M.), the Evan Dunbar Foundation (J.M.M.), the Rally Foundation (J.M.M.), Andrew's Army Foundation (J.M.M.), the Abramson Family Cancer Research Institute (J.M.M.), a Howard Hughes Medical Institute Research Training Fellowship (K.B.), a fellowship

from Associazione Oncologia Pediatrica e Neuroblastoma (M.C.), a Research Development Award from the Cotswold Foundation (H.H.), UL1-RR024134-03 (H.H.) and an Institutional Development Award to the Center for Applied Genomics from the Children's Hospital of Philadelphia (H.H.).

Author Contributions H.H. and J.M.M. conceived the study, guided interpretation of results and helped preparation of the manuscript. K.W., H.Z. and C.H. performed SNP association analysis. K.W., S.J.D., E.F.A. and J.J. performed gene expression and copy number analysis. C.W. and K.B. performed PCR validation of gene expression data. C.W., R.W.S., K.B., P.A.M., S.J.D. and K.A.C. performed and/or analysed shRNA transfection and *LMO1* overexpression experiments. N.S. and H.S. generated viral construct for human *LMO1* complementary DNA. M.C. and A.I. performed the replication study on the Italian case series, and N.R. performed the replication study on the UK case series. P.W.M. and W.B.L. performed outcome and clinical covariate analyses on the Children's Oncology Group samples. C.H., C.K., E.F., M.G., W.G. and R.C. generated the genotyping data. L.N. and M.D. helped with data analysis. S.F.A.G., Y.P.M., H.L. and M.D. advised on data interpretation. K.W. drafted the manuscript; H.H., J.M.M. and other authors edited it.

Author Information Microarray data are deposited in the GEO database under accession number GSE3960. The genotypic and phenotypic information from this study is deposited in dbGaP (www.ncbi.nlm.gov/gap) under accession number phs000124.v2.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.H. (hakonarson@chop.edu) or J.M.M. (maris@chop.edu).

A role for mitochondria in NLRP3 inflammasome activation

Rongbin Zhou¹, Amir S. Yazdi¹, Philippe Menu¹ & Jürg Tschopp¹

An inflammatory response initiated by the NLRP3 inflammasome is triggered by a variety of situations of host ‘danger’, including infection and metabolic dysregulation^{1,2}. Previous studies suggested that NLRP3 inflammasome activity is negatively regulated by autophagy and positively regulated by reactive oxygen species (ROS) derived from an uncharacterized organelle. Here we show that mitophagy/autophagy blockade leads to the accumulation of damaged, ROS-generating mitochondria, and this in turn activates the NLRP3 inflammasome. Resting NLRP3 localizes to endoplasmic reticulum structures, whereas on inflammasome activation both NLRP3 and its adaptor ASC redistribute to the perinuclear space where they co-localize with endoplasmic reticulum and mitochondria organelle clusters. Notably, both ROS generation and inflammasome activation are suppressed when mitochondrial activity is dysregulated by inhibition of the voltage-dependent anion channel. This indicates that NLRP3 inflammasome senses mitochondrial dysfunction and may explain the frequent association of mitochondrial damage with inflammatory diseases.

The NLRP3 inflammasome is a molecular platform activated upon signs of cellular ‘danger’ to trigger innate immune defences through the maturation of pro-inflammatory cytokines such as interleukin (IL)-1 β ¹. Strong associations of a number of human heritable and acquired diseases with dysregulated inflammasome activity highlight the importance of the NLRP inflammasome in regulating immune responses³. Key components of a functional NLRP3 inflammasome are NLRP3, the adaptor protein ASC and caspase-1 (ref. 1). Upon detecting cellular stress, NLRP3 recruits ASC and procaspase-1, which results in caspase-1 activation and processing of cytoplasmic targets, including the pro-inflammatory cytokines IL-1 β and IL-18.

A wide variety of danger signals activate the NLRP3 inflammasome. These include pathogen-associated molecular patterns⁴ and host-derived molecules that are indicative of cellular damage (danger-associated molecular patterns) such as uric-acid crystals¹. The mechanisms by which these structurally distinct molecules trigger NLRP3 inflammasome activation are currently unclear. One of the models proposes that NLRP3 is activated by a common pathway of ROS⁵. The source of ROS is currently unclear, but we previously suggested the involvement of one or several of the seven known NADPH oxidases⁵. However, macrophages deficient in subunits specific to three of the seven NADPH oxidases, that is, NOX1, NOX2 and NOX4, respond normally to inflammasome activators (C. Dostert and J.T., unpublished observations) or have even slightly increased activity⁶, which suggests a possible role of other NADPH oxidase members or functional redundancy. Alternatively, further ROS required for inflammasome activation could be generated by other cellular sources.

The main source of cellular ROS is mitochondria. Various stress conditions, including increased metabolic rates, hypoxia, or membrane damage all markedly induce mitochondrial ROS production⁷. To investigate a possible implication of mitochondria in inflammasome activation, we artificially induced ROS production in mitochondria by blocking key enzymes of the respiratory chain. Complex I is one of the main sites at which electrons can leak to oxygen and result in superoxide

production⁷. In agreement with previous reports⁸, addition of the complex I inhibitor rotenone resulted in the loss of mitochondrial membrane potential and robust ROS production (Fig. 1a, b). This was determined using three types of mitochondria-specific labels that distinguish respiring (Mitotracker deep red), total (Mitotracker green) and ROS-generating mitochondria (MitoSOX) (Fig. 1a, b). ROS-generating mitochondria were also observed when complex III was inhibited by antimycin A, whereas the complex II inhibitor thenoyltrifluoroacetone (TTFA) had only a minor effect (Fig. 1a, b).

A correlation between mitochondrial ROS activity and the presence of active IL-1 β in the supernatant of the human THP1 macrophage cell line was observed. Whereas rotenone- or antimycin-treated cells showed increased IL-1 β secretion in a dose-dependent manner, no or little cytokine was detectable if cells were treated with TTFA (Fig. 1c). In THP1 cells with knocked-down NLRP3 or caspase-1, or in bone-marrow-derived macrophages (BMDMs) from *Nlrp3*^{-/-} mice, the respiratory chain inhibitors did not induce IL-1 β or caspase-1 secretion (Fig. 1d and Supplementary Fig. 1a). In contrast, macrophages lacking the IPAF/NLRC4 inflammasome responded in a similar way to wild-type cells, indicating that activation of caspase-1 and processing and secretion of IL-1 β by inhibitors of the respiratory chain were mediated by the NLRP3 inflammasome but not the IPAF inflammasome. As previously observed for MSU, activation of the NLRP3 inflammasome by rotenone and antimycin was blocked by the ROS inhibitor APDC (Supplementary Fig. 1b). Although inhibition of complex I and III activity leads to robust ROS production, damage of the mitochondria and decrease of membrane potential appeared to be partial and did not result in cell death (Supplementary Fig. 2). Consistent with this, complete breakdown of the mitochondrial membrane potential induced by high doses of the respiratory chain uncoupler CCCP did cause only little ROS production and IL-1 β release. Nonetheless, CCCP used at a lower concentration that only partially impairs the membrane potential also spontaneously led to NLRP3 inflammasome activation (Supplementary Fig. 2). These data are compatible with the notion that ROS generated by mitochondria having reduced membrane potential can lead to NLRP3 inflammasome activation.

To avoid cellular damage, ROS-generating mitochondria are constantly removed by mitophagy, a specialized process of autophagy⁹. In agreement with this established mechanism of mitochondrial disposal, autophagy-associated LC3 puncta were found to accumulate around mitochondria after the addition of rotenone (Fig. 2a). We therefore speculated that inhibition of mitophagy/autophagy should lead to the accumulation of ROS-producing damaged mitochondria, and as a consequence, to the activation of the inflammasome. To this end, the mitophagy/autophagy inhibitor 3-methyladenine (3-MA) was added to THP1 macrophages, which, as expected, resulted in the accumulation of damaged mitochondria and increased concentrations of mitochondrial ROS (Fig. 2b and Supplementary Fig. 3a). Similar to mitochondrial respiratory chain inhibition, 3-MA-induced ROS generation was paralleled by the dose-dependent secretion of IL-1 β within 6 h of mitophagy/autophagy inhibition (Supplementary Fig. 3b), which was blocked by the antioxidant APDC (Supplementary Fig. 3c). Processing

¹Department of Biochemistry, Center of Immunity and Infection, University of Lausanne, Chemin des Boveresses 155, CH-1066 Epalinges, Switzerland.

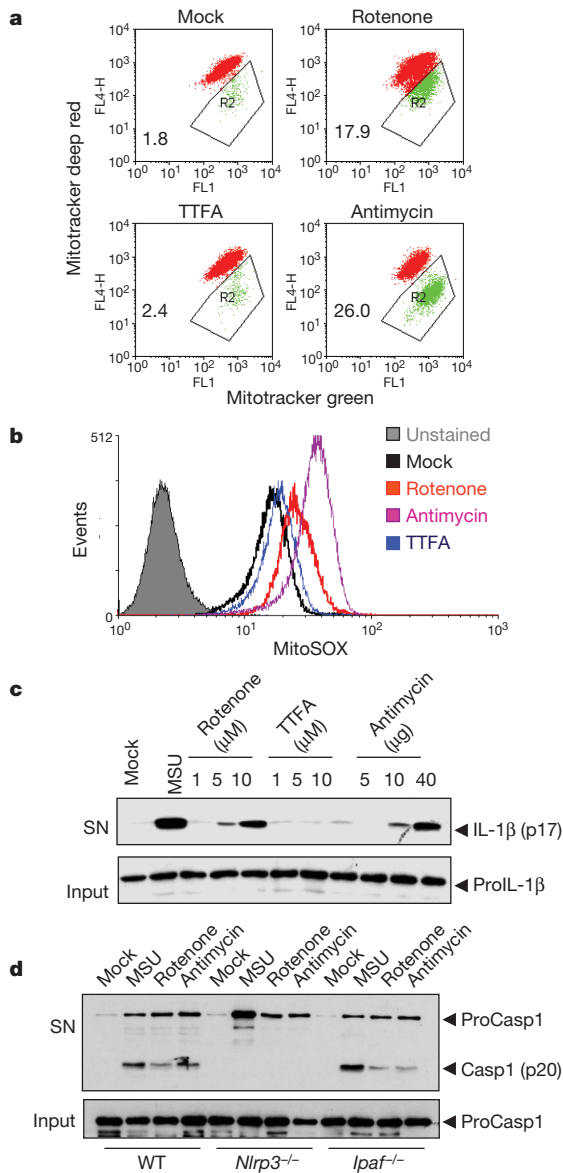


Figure 1 | Mitochondrial ROS can trigger NLRP3 inflammasome activation.

a, b, THP1 cells were stimulated with rotenone (10 μM), TTFA (10 μM) or antimycin (40 $\mu\text{g ml}^{-1}$) for 6 h and then stained with Mitotracker green and Mitotracker deep red (**a**) or MitoSOX (**b**) for 30 min and analysed by flow cytometry. **c,** THP1 cells were stimulated for 6 h with the indicated amounts of rotenone, TTFA or antimycin. Supernatants (SN) and cell extracts (input) were analysed by western blotting as indicated. **d,** LPS-primed bone-marrow-derived macrophages (BMDMs) from wild-type (WT), *Nlrp3* or *Ipaf* deficient mice were stimulated for 6 h with MSU (150 $\mu\text{g ml}^{-1}$), rotenone (10 μM for THP1 cells and 40 μM for BMDMs) or antimycin (40 $\mu\text{g ml}^{-1}$ for THP1 cells and 10 $\mu\text{g ml}^{-1}$ for BMDMs). The release of caspase-1 (western blot) was then determined. Data shown are representative of three independent experiments.

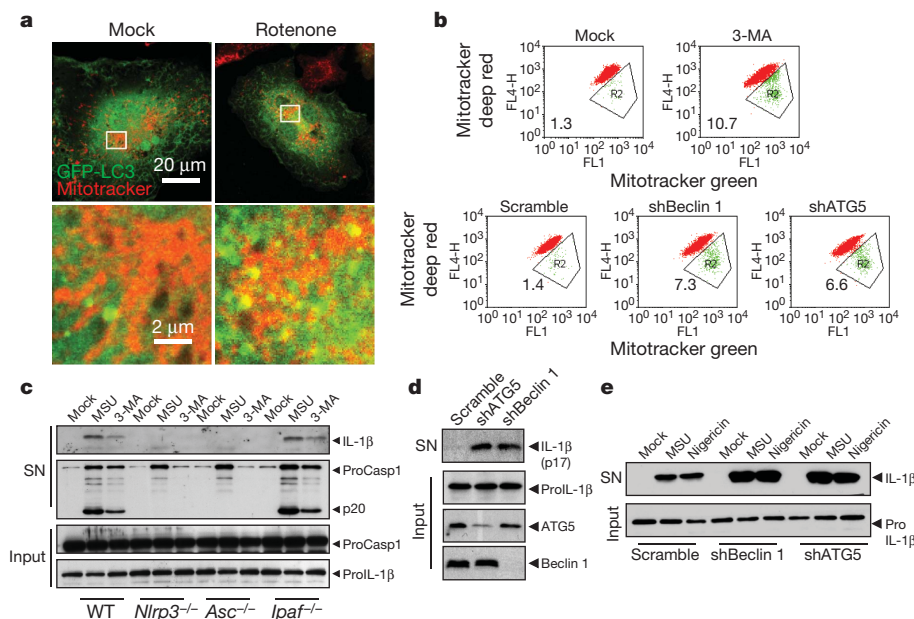
of proIL-1 β caused by the blockade of mitophagy/autophagy was NLRP3- and ASC-dependent and not reliant on the IPAF inflammasome (Fig. 2d).

Because most pharmacological drugs have off-target effects, we next inhibited mitophagy/autophagy specifically through downregulation of two proteins that have crucial functions in mitophagy/autophagy, that is, beclin 1, a Bcl-2- and PI3KIII-interacting protein, and ATG5, a protein essential for autophagosome formation¹⁰. In contrast to the pharmacological inhibitors, spontaneous inflammasome activation was less efficient, as cells with knocked down beclin 1 and ATG5 had to be cultured for 24 h before we could detect active IL-1 β in the culture supernatant (Fig. 2e, f). Inhibition of mitophagy/autophagy through downregulation of beclin 1 or ATG5 also sensitized macrophages for the release of IL-1 β induced by MSU or nigericin (Fig. 2f). This augmented response was not associated with increased cell death (Supplementary Fig. 3d). It is therefore feasible that a functional mitophagy/autophagy system acts as a scavenger of mitochondrial ROS through removal of damaged mitochondria and thereby suppresses NLRP3 inflammasome activation.

ROS are short-lived and can act as a signalling messenger only for a short distance¹¹. Thus, NLRP3 should ideally be localized in close vicinity to mitochondria, allowing efficient sensing of the presence of damaged ROS-generating mitochondria. Of the 22 NLR members, only one, NOD5 (also called NLRX1), has a predicted mitochondrial import sequence and thus resides in the mitochondria^{12,13}. Using confocal microscopy, the localization of NLRP3 in THP1 macrophages was examined. Under non-stimulatory conditions, most NLRP3 protein was found to localize to cytoplasmic granular structures (Fig. 3). Further analysis showed a significant overlap of NLRP3 with the

Figure 2 | Inhibition of autophagy/mitophagy results in ROS generation and inflammasome activation.

a, BMDMs expressing GFP-LC3 were stimulated with rotenone (40 μM) for 3 h and the co-localization of mitochondria and GFP-LC3 dots were analysed using confocal microscopy. **b,** THP1 cells stimulated with 3-methyladenine (3-MA, 10 mM) for 24 h or THP1 cells stably expressing shRNA against beclin 1 or ATG5 were stained with Mitotracker green and Mitotracker deep red for 30 min, and analysed by flow cytometry. **c,** LPS-primed BMDMs from wild-type, *Nlrp3*^{-/-}, *Asc*^{-/-} or *Ipaf*^{-/-} knockout mice were stimulated with MSU (150 $\mu\text{g ml}^{-1}$) or 3-MA (10 mM) for 6 h and the release of active caspase-1 and IL-1 β was determined. **d,** THP1 cells stably expressing shRNA against beclin 1 or ATG5 were incubated for 24 h and media supernatants (SN) and inputs were analysed by western blotting. **e,** THP1 cells stably expressing shRNA against beclin 1 or ATG5 were stimulated with MSU or nigericin for 6 h. Data shown are representative of three independent experiments.



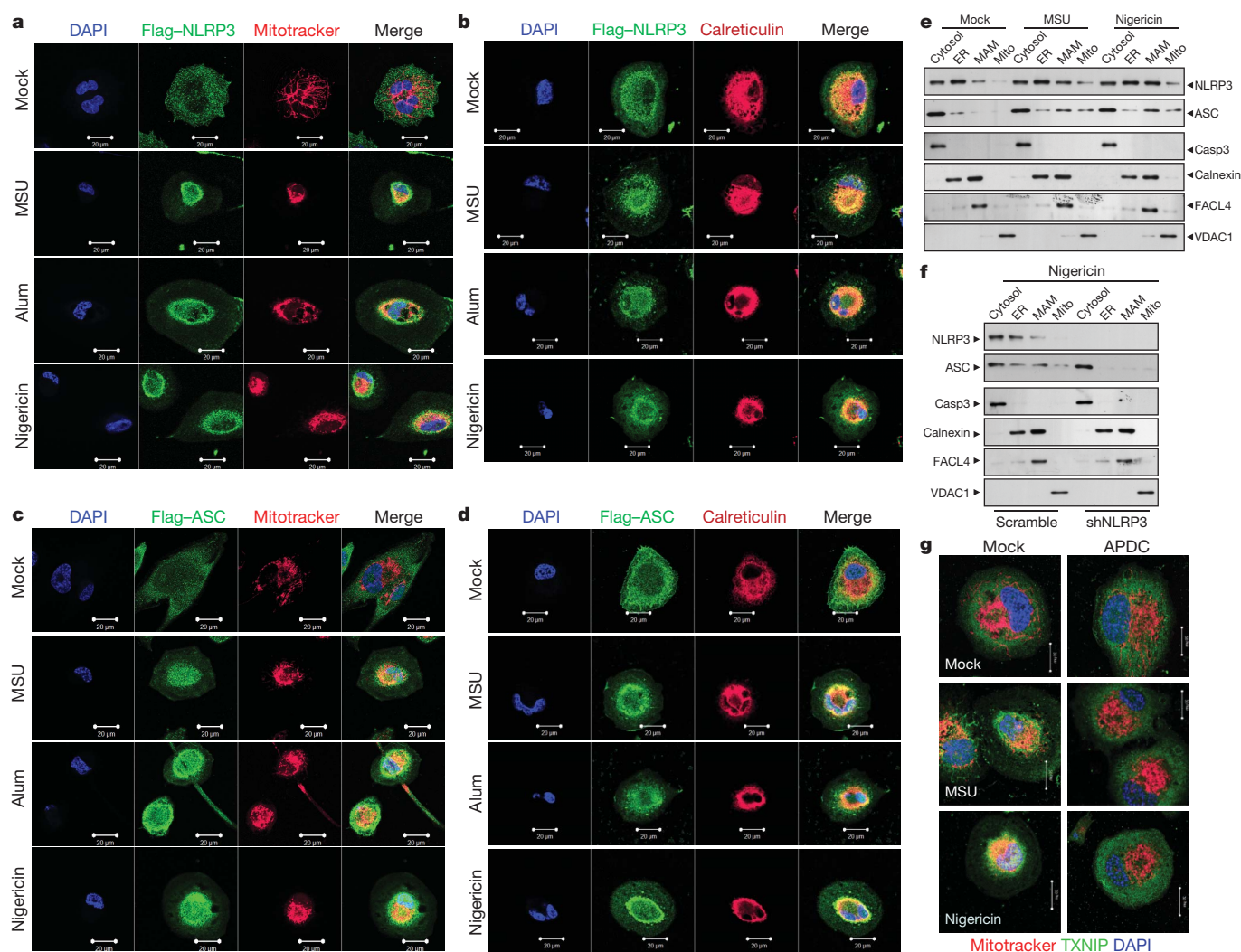


Figure 3 | Co-localization of the NLRP3 inflammasome components and endoplasmic reticulum-mitochondria. **a–d**, THP1 cells expressing Flag-NLRP3 (**a**, **b**) or Flag-ASC (**c**, **d**) were analysed for the co-localization of Flag-NLRP3 with the endoplasmic reticulum (ER) marker calreticulin or mitochondria (Mitotracker) using confocal microscopy. **e**, THP1 cells were stimulated with MSU or nigericin and then fractionated. The cytosolic, ER, MAM and mitochondrial fractions were analysed by western blotting as

indicated in the text. **f**, THP1 cells stably expressing shRNA against NLRP3 were stimulated with nigericin and subcellular fractions analysed as in **e**. **g**, TXNIP associates with mitochondria after NLRP3 inflammasome activation in a ROS-dependent manner. THP1 cells were stimulated with MSU and nigericin, and TXNIP localization was investigated in the presence or absence of the antioxidant APDC. Data shown are representative of two or three independent experiments. Scale bars: **a–g**, 20 μ m.

endoplasmic reticulum (ER) marker calreticulin, whereas no or very little co-localization was detected with Mitotracker, the Golgi marker Giantin or the lysosome/phagosome staining dye lysotracker (Fig. 3a, b and Supplementary Fig. 4). Primarily ER and not mitochondrial staining of NLRP3 was also detected when cells were analysed by electron microscopy (Supplementary Fig. 5). This localization changed significantly after inflammasome stimulation with MSU, alum or nigericin. NLRP3 relocated into the perinuclear space and co-localized with structures that stained positively for both the ER and mitochondria (Fig. 3a, b and Supplementary Fig. 6). A similar perinuclear, ER/mitochondria localization was detected for ASC after NLRP3 activation, although in contrast to NLRP3 most ASC was observed in the cytoplasm in resting cells (Fig. 3c, d and Supplementary Figs 4 and 6).

To confirm the NLRP3 association with the ER/mitochondria observed by microscopy, subcellular fractionation studies were performed. With this approach, NLRP3 was detected in both ER and cytoplasmic fractions in a roughly similar amount (Fig. 3e). ER membranes can tightly associate with mitochondria forming mitochondria-associated ER membranes (MAMs). MAMs are important, among other functions, in transferring lipids and Ca^{2+} from the ER to the mitochondria¹⁴. Addition of MSU and, in particular, nigericin induced

NLRP3 to associate partly with MAMs (Fig. 3e). In contrast to NLRP3, the majority of the ASC was found in the cytoplasmic fraction before inflammasome activation, and only a very minor portion of ASC was associated with the ER fraction. On stimulation with MSU or nigericin, however, the amount of ASC found in the mitochondrial fraction and particularly in the MAMs fraction increased considerably. ASC translocation to the MAMs and mitochondria was NLRP3 dependent, as no or little translocation was found in nigericin-activated THP1 cells in which NLRP3 was downregulated by short hairpin (sh)RNA (Fig. 3f).

We have previously reported that TXNIP, a protein implicated in type 2 diabetes, associated with NLRP3 in a ROS-dependent manner after its detachment from thioredoxin¹⁵. We therefore anticipated that TXNIP translocates to MAMs/mitochondria on NLRP3 inflammasome activation. Consistent with this idea, MSU and nigericin induced TXNIP to redistribute to mitochondria in a ROS-dependent manner (Fig. 3g), in agreement with a recently published report¹⁶.

Although the above results indicated that the prolonged presence of damaged and ROS-producing mitochondria might be implicated in inflammasome activation, the evidence was still indirect. We therefore sought to obtain more direct proof by inhibiting the activity of voltage-dependent anion channels (VDAC), which are the most abundant

proteins of the outer mitochondrial membrane and the major channels for the exchange of metabolites and ions between the mitochondria and other cellular compartments including the ER¹⁷. As such, the three VDAC isoforms are important regulators of mitochondrial metabolic activity, which is ultimately required for ROS production. We down-regulated expression of each of the three human VDAC isoforms by shRNA (Supplementary Fig. 7a) and measured inflammasome activation by MSU, R837, silica, alum and nigericin (Fig. 4a). In cells with knocked-down VDAC1, caspase-1 activation and IL-1 β secretion were considerably impaired for all inflammasome activators examined. As a consequence of VDAC1 knockdown, mitochondrial ROS was highly diminished (Supplementary Fig. 8). In contrast to the NLRP3 inflammasome, VDAC1 was not essential for the activation of the IPAF or the AIM2 inflammasome (Supplementary Fig. 7b). A significant reduction of IL-1 β secretion was also seen in cells where VDAC2 was knocked down (Fig. 4b). In all cases, impairment was seen with two distinct shRNA constructs (Supplementary Fig. 7). In contrast to VDAC1 and VDAC2, downregulation of VDAC3 had no effect on NLRP3 inflammasome activity (Supplementary Fig. 7c). The absence of VDAC1 also inhibited MSU- or nigericin-induced NLRP3 translocation to perinuclear areas (Fig. 4c).

VDAC activity is regulated by hexokinase and Bcl-2 family members¹⁷. Overexpression of Bcl-2 leads to partial VDAC closure and a concomitant decrease of mitochondrial Ca^{2+} levels and ROS production. We therefore examined whether increased Bcl-2 levels would have an impact on NLRP3 inflammasome activity and found that in supernatants of MSU-, alum- or nigericin-stimulated macrophages isolated from Bcl-2-overexpressing transgenic mice, IL-1 β levels were decreased when compared to cells from wild-type mice (Supplementary Fig. 9a). In contrast, no influence of Bcl-2 was detectable on *Salmonella*-mediated IPAF-inflammasome activity (Supplementary Fig. 9a) or LPS-mediated tumour-necrosis factor (TNF) secretion (Supplementary Fig. 9b).

It is now widely recognized that, apart from bioenergetic ATP production, mitochondria also have a crucial role in cell signalling events such as apoptotic cell death. Apoptotic signal transmission to the mitochondria results in the efflux of a number of potential apoptotic regulators such as cytochrome *c* to the cytosol that trigger caspase activation and lead to cell death. Our results now provide evidence for an unanticipated additional role of mitochondria, namely the orchestration of the inflammatory response.

Mitochondria are the major source of cellular ROS and it is therefore not completely unexpected that we found mitochondrial ROS capable

of NLRP3 inflammasome activation. This notion is based on the following observations: (1) inhibition of complex I or III of the mitochondrial respiratory chain, known to result in ROS generation, causes unprompted NLRP3 inflammasome activation; (2) inhibition of mitophagy/autophagy, resulting in the prolonged presence of damaged, ROS-generating mitochondria, leads to spontaneous inflammasome activation; (3) NLRP3 and ASC co-localize with mitochondria and MAMs in the presence of NLRP3 inflammasome activators; (4) knock down by shRNA or inhibition by Bcl-2 of VDAC ion channels that are crucial in mitochondrial activity and ROS generation significantly impair NLRP3 inflammasome activation. The dependence on VDAC is specific for the NLRP3 inflammasome and is not seen with the AIM2 or IPAF inflammasomes, indicating that the signalling pathways leading to their activation are distinct.

VDAC is essential for the uptake of Ca^{2+} into the mitochondria from MAMs to promote mitochondrial metabolic activity¹⁷. Upon NLRP3 inflammasome activation, a substantial portion of the inflammasome is associated to MAMs, indicating that NLRP3 checks and modulates mitochondrial activity.

There is a large amount of literature proposing a link between mitochondrial malfunction, ROS and chronic inflammatory diseases. Damage to mitochondria is now understood to have a role in the pathogenesis of a wide range of seemingly unrelated disorders¹⁸. For example, mutations in PINK1, PARKIN and DJ-1 are frequent causes of recessive Parkinson's disease. Both PINK1 and PARKIN are crucial in the removal of damaged mitochondria by mitophagy, whereas DJ-1 is localized to mitochondria and has a role in oxidative stress protection. Mutations in proteins involved in mitophagy/autophagy, such as ATG16 and IRGM, are also frequently found in Crohn's disease¹⁹. In line with this, increased IL-1 β production is observed in mice deficient in the autophagy gene *ATG16* (ref. 20). It is, however, unlikely that inhibition of mitophagy is generally used as a means to activate the NLRP3 inflammasome signalling pathway. In our opinion, a more direct mechanism, that is, an excessive induction of mitochondrial oxidative phosphorylation, is more likely.

Taken together, our data unravel unexpected mechanistic parallels between signalling pathways leading to apoptosis and inflammation. In both instances, signals converge at the level of mitochondria where VDAC activity seems to have a role for the activation of two seemingly unrelated downstream processes. VDAC is not only crucial for inflammasome activation, as shown here, but also has a role in apoptosis induction by Bax²¹. The mechanism that determines whether mitochondria

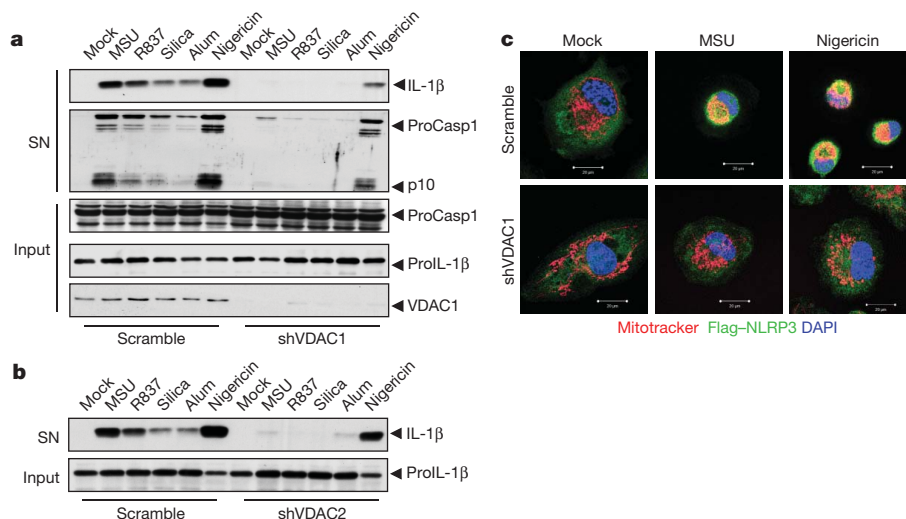


Figure 4 | VDAC is essential for NLRP3 inflammasome activation. **a**, THP1 cells stably expressing shVDAC1 were stimulated with various NLRP3 inflammasome activators and then analysed for caspase-1 activation and IL-1 β secretion. **b**, THP1 cells expressing shRNA against VDAC2 were stimulated

with MSU, R837, silica, alum or nigericin and analysed for IL-1 β secretion. **c**, THP1 cells expressing shRNA against VDAC1 were stimulated with MSU and nigericin, and NLRP3 localization was analysed by confocal microscopy. Scale bars: 20 μm .

induce inflammasome or apoptosome assembly is currently unclear. Whereas ROS is essential for the activation of both the inflammasome and apoptosome²², apoptosome formation requires in addition cytochrome *c* release from mitochondria into the cytosol. There is no evidence for cytochrome *c* being implicated in inflammasome activation, yet it cannot be ruled out that recognition of a different protein released from partially damaged mitochondria is required. The decision to induce inflammation or apoptosis must be tightly controlled to avoid pathological conditions associated with a chronic inflammatory response. Regrettably, this control mechanism seems to be easily deranged, as shown by the many diseases that are associated with mutations in genes that affect mitochondrial function.

METHODS SUMMARY

Mice. *Nlrp3*^{-/-}, *Asc*^{-/-}, *Ipa1*^{-/-} and H2K-Bcl-2 transgenic mice were described previously^{23–25}. All mice were on a C57BL/6 background except for experiments used for NLRP1b inflammasome activation (BALB/c).

Cell fractionation. MAMs, mitochondria and microsomes were isolated from PMA-differentiated THP1 cells as previously described²⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 June; accepted 11 November 2010.

Published online 1 December 2010.

- Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
- Latz, E. The inflammasomes: mechanisms of activation and function. *Curr. Opin. Immunol.* **22**, 28–33 (2010).
- Kastner, D. L., Aksentijevich, I. & Goldbach-Mansky, R. Autoinflammatory disease reloaded: a clinical perspective. *Cell* **140**, 784–790 (2010).
- Franchi, L., Munoz-Planillo, R., Reimer, T., Eigenbrod, T. & Nunez, G. Inflammasomes as microbial sensors. *Eur. J. Immunol.* **40**, 611–615 (2010).
- Dostert, C. *et al.* Innate immune activation through Nalp3 inflammasome sensing of asbestos and silica. *Science* **320**, 674–677 (2008).
- Latz, E. NOX-free inflammasome activation. *Blood* **116**, 1393–1394 (2010).
- Brookes, P. S., Yoon, Y., Robotham, J. L., Anders, M. W. & Sheu, S.-S. Calcium, ATP, and ROS: a mitochondrial love-hate triangle. *Am. J. Physiol. Cell Physiol.* **287**, C817–C833 (2004).
- Li, N. *et al.* Mitochondrial complex I inhibitor rotenone induces apoptosis through enhancing mitochondrial reactive oxygen species production. *J. Biol. Chem.* **278**, 8516–8525 (2003).
- Goldman, S. J., Taylor, R., Zhang, Y. & Jin, S. Autophagy and the degradation of mitochondria. *Mitochondrion* **10**, 309–315 (2010).
- Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).
- Veal, E. A., Day, A. M. & Morgan, B. A. Hydrogen peroxide sensing and signaling. *Mol. Cell* **26**, 1–14 (2007).
- Moore, C. B. *et al.* NLRX1 is a regulator of mitochondrial antiviral immunity. *Nature* **451**, 573–577 (2008).
- Tattoli, I. *et al.* NLRX1 is a mitochondrial NOD-like receptor that amplifies NF- κ B and JNK pathways by inducing reactive oxygen species production. *EMBO Rep.* **9**, 293–300 (2008).
- Hayashi, T., Rizzuto, R., Hajnoczky, G. & Su, T. P. MAM: more than just a housekeeper. *Trends Cell Biol.* **19**, 81–88 (2009).
- Zhou, R., Tardivel, A., Thorens, B., Choi, I. & Tschopp, J. Thioredoxin-interacting protein links oxidative stress to inflammasome activation. *Nature Immunol.* **11**, 136–140 (2010).
- Saxena, G., Chen, J. & Shalev, A. Intracellular shuttling and mitochondrial function of thioredoxin-interacting protein. *J. Biol. Chem.* **285**, 3997–4005 (2010).
- Colombini, M. VDAC: the channel at the interface between mitochondria and the cytosol. *Mol. Cell. Biochem.* **256**, 107–115 (2004).
- Oliveira, J. M. Nature and cause of mitochondrial dysfunction in Huntington's disease: focusing on Huntingtin and the striatum. *J. Neurochem.* **114**, 1–12 (2010).
- Restivo, N. L., Srivastava, M. D., Schafer, I. A. & Hoppel, C. L. Mitochondrial dysfunction in a patient with crohn disease: possible role in pathogenesis. *J. Pediatr. Gastroenterol. Nutr.* **38**, 534–538 (2004).
- Saitoh, T. *et al.* Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1 β production. *Nature* **456**, 264–268 (2008).
- Yamagata, H. *et al.* Requirement of voltage-dependent anion channel 2 for pro-apoptotic activity of Bax. *Oncogene* **28**, 3563–3572 (2009).
- Sato, T. *et al.* Fas-mediated apoptosome formation is dependent on reactive oxygen species derived from mitochondrial permeability transition in Jurkat cells. *J. Immunol.* **173**, 285–296 (2004).
- Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
- Martinon, F., Petrilli, V., Mayor, A., Tardivel, A. & Tschopp, J. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440**, 237–241 (2006).
- Domen, J., Cheshier, S. H. & Weissman, I. L. The role of apoptosis in the regulation of hematopoietic stem cells: Overexpression of Bcl-2 increases both their number and repopulation potential. *J. Exp. Med.* **191**, 253–264 (2000).
- Wieckowski, M. R., Giorgi, C., Lebiedzinska, M., Duszynski, J. & Pinton, P. Isolation of mitochondria-associated membranes and mitochondria from animal tissues and cells. *Nature Protocols* **4**, 1582–1590 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by grants of the Swiss National Science Foundation and by the Institute for Arthritis Research. A.S.Y. is a recipient of a stipend of the DFG. We would like to thank K. Schroder, C. Thomas and J. Vince for critical reading of the manuscript and G. W. Knott, EPFL, Lausanne, for help in collecting electron micrographs.

Author Contributions R.Z., A.S.Y. and P.M. devised and performed the experiments. J.T. supervised the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.T. (jurg.tschopp@unil.ch).

METHODS

Mice. *Nlrp3*^{-/-}, *Asc*^{-/-}, *Irf1*^{-/-} and H2K-Bcl-2 transgenic mice were described previously^{23–25}. All mice were on a C57BL/6 background except for experiments used for NLRP1b inflammasome activation (BALB/c).

Cell fractionation. MAMs, mitochondria and microsomes were isolated from PMA-differentiated THP1 cells as previously described²⁶.

Reagents. Nigericin, rotenone, antimycin, thenoyltrifluoroacetone (TTFA), uric acid, (2R,4R)-4-aminopyrrolidine-2, 4-dicarboxylate (APDC), 3-methyladenine (3-MA), ATP, menadione, carbonyl cyanide 3-chlorophenylhydrazone (CCCP), poly(AT) and silica were purchased from Sigma. MitoSOX, Mitotracker green, Mitotracker deep red, JC-1 and lysotracker red were from Invitrogen. LDH release kits were from Roche. Ultrapure LPS and R837 were obtained from InvivoGen. Inject alum was from Pierce. *Salmonella typhimurium* is a gift from R. Van Bruggen. Anti-human proIL-1 β was produced in house. The anti-NLRP3 antibodies (Cryo-2) were from Adipogen. Anti-human cleaved IL-1 β (D116), anti-ATG5 (2630S), anti-beclin 1 (3738), anti-calnexin (2433S), anti-caspase-3 (9662) and anti-VDAC1 (4866) antibodies were purchased from Cell Signaling. Anti-ASC antibody (AL177) was from Adipogen. Anti-TXNIP (40-3700) antibody was from Invitrogen. Anti-VDAC3 antibody (sc-79341), anti-FACL-4 (sc-47997) and anti-human caspase-1 (sc-622) were from Santa Cruz. Anti-VDAC2 antibody (ab77160), anti-Giantin (ab24586), anti-GAPDH (ab8245) and anti-calreticulin (ab14234) were from Abcam. Anti-mouse IL-1 β antibody was a gift from R. Solari. Anti-mouse caspase-1 (p20) antibody was a gift from P. Vandenabeele. Anti-Flag (M2) antibody was from List Biological Laboratories. All tissue culture reagents were from Invitrogen.

Generation of stable THP1 cells expressing Flag-NLRP3 or Flag-ASC. Retroviral vector (LZRS-MS-IRES-ZEO/PBR-Flag-NLRP3 or Flag-ASC) was co-transfected with the helper plasmids VSV-G and Hit60 into 293T cells by calcium phosphate transfection. Culture supernatants containing recombinant viral particles were harvested and used to infect THP1 cells in the presence of polybrene (8 mg ml⁻¹). To establish stable cell lines, THP1 cells were selected with zeocin (1 mg ml⁻¹) on day 3 after infection.

Transduction of GFP-LC3 in BMDMs. Lentiviral vector (plex-GFP-LC3) was co-transfected with the helper plasmids pCMV Δ R8.91 and pMDG into 293T cells by calcium phosphate transfection. Culture supernatants containing recombinant viral particles were harvested and used to infect BMDMs on day 3 after differentiation. On day 7, the cells can be used for advanced experiments.

Cell preparation and stimulation. Human THP1 cells were cultured in RPMI 1640 media, supplemented with 10% FBS. THP1 cells were differentiated for 3 h with 100 nM phorbol-12-myristate-13-acetate (PMA). Bone marrow macrophages were derived from tibia and femoral bone marrow progenitors as described²⁷, and were primed for 4 h with 100 ng ml⁻¹ Ultra-pure LPS. For the induction of inflammasome activation, 10⁶ LPS-primed bone marrow macrophages or PMA-differentiated THP1 cells plated in 12-well plates were treated with MSU (150 μ g ml⁻¹), alum (200 μ g ml⁻¹), silica (200 μ g ml⁻¹), R837 (15 μ g ml⁻¹) for 6 h or with ATP (5 mM) or nigericin (15 μ M) for 30 min. For poly(dA:dT) transfection,

poly(dA:dT) was transfected using Lipofectamine (4 μ l ml⁻¹) as per the manufacturer's protocol (Invitrogen). Cell extracts and precipitated supernatants were analysed by immunoblot.

Generation of stable THP1 cells expressing shRNA. THP1 cells stably expressing shRNA were obtained as previously described²⁸; shRNA against NLRP3 and caspase-1 plasmids has been described¹⁵ and shRNA plasmids against beclin-1, ATG5, VDAC1, VDAC2 and VDAC3 were from Sigma.

Confocal microscopy. PMA-differentiated THP1 cells were plated on coverslips for 3 days and then used for stimulation or staining with Mitotracker red (50 nM) or Lysotracker (200 nM). After washing two times with PBST, the cells were fixed with PFA 4% in PBS for 15 min and then washed three times with PBST. After permeabilization with Triton X-100 and blocking with 10% goat serum in PBS, cells were incubated with primary antibodies (in 10% goat serum) overnight at 4 °C. After washing with PBST, cells were incubated with secondary antibodies (Invitrogen) in 10% goat serum-PBS for 60 min and rinsed in PBST. Confocal microscopy analyses were carried out using a Zeiss LSM510.

Flow cytometric analyses. Mitochondrial mass was measured by fluorescence levels upon staining with Mitotracker green and Mitotracker deep red at 50 nM for 30 min at 37 °C. Mitochondria-associated ROS levels were measured by staining cells with MitoSOX at 2.5 μ M for 30 min at 37 °C. Mitochondria membrane potential was measured using the kit from Invitrogen and performed according to the manufacturer's instructions. Cells were then washed with PBS solution and resuspended in cold PBS solution containing 1% FBS for FACS analysis.

Salmonella infection. *Salmonella typhimurium* was pre-cultured on day 1. On day 2, differentiated THP1 cells were infected for 1 h with the *S. typhimurium* culture (1:100). Cells were washed and then incubated for 3 h in OptiMEM medium with gentamycin (Invitrogen).

ELISA. Cell culture supernatants were assayed for mouse IL-1 β and TNF (R&D) according to manufacturer's instructions.

Immunoelectron microscopy. For the pre-embedding immuno-EM of cells, PMA-differentiated Flag-Nlrp3-THP1 cells were fixed with 4% PFA for 15 min and were cryoprotected in 2% glycerol and 20% DMSO in PBS. After two freeze-thaw cycles, the cells were blocked in PBS with 0.2% BSA and 0.02% Triton X-100 for 1 h. After washing, the cells were incubated with anti-Flag antibody and gold-conjugated secondary antibody. Silver enhancement was performed by incubating cells with silver enhancement reagent for 1 h in room temperature. After osmification in 0.5% osmium tetroxide for 15 min and dehydration in graded alcohol, the samples were embedded in Durcupan.

Statistical analyses. All values were expressed as the mean \pm s.e.m. of individual samples. Samples were analysed using Student's *t*-test for two groups and ANOVA for multiple groups.

27. Didierlaurent, A. *et al.* Tollip regulates proinflammatory responses to interleukin-1 and lipopolysaccharide. *Mol. Cell. Biol.* **26**, 735–742 (2006).
28. Papin, S. *et al.* The SPRY domain of Pyrin, mutated in familial Mediterranean fever patients, interacts with inflammasome components and inhibits proIL-1 β processing. *Cell Death Differ.* **14**, 1457–1466 (2007).

Hydrostatic pressure and the actomyosin cortex drive mitotic cell rounding

Martin P. Stewart^{1,2}, Jonne Helenius¹, Yusuke Toyoda³, Subramanian P. Ramanathan¹, Daniel J. Muller¹ & Anthony A. Hyman³

During mitosis, adherent animal cells undergo a drastic shape change, from essentially flat to round^{1–3}. Mitotic cell rounding is thought to facilitate organization within the mitotic cell and be necessary for the geometric requirements of division^{4–7}. However, the forces that drive this shape change remain poorly understood in the presence of external impediments, such as a tissue environment². Here we use cantilevers to track cell rounding force and volume. We show that cells have an outward rounding force, which increases as cells enter mitosis. We find that this mitotic rounding force depends both on the actomyosin cytoskeleton and the cells' ability to regulate osmolarity. The rounding force itself is generated by an osmotic pressure. However, the actomyosin cortex is required to maintain this rounding force against external impediments. Instantaneous disruption of the actomyosin cortex leads to volume increase, and stimulation of actomyosin contraction leads to volume decrease. These results show that in cells, osmotic pressure is balanced by inwardly directed actomyosin cortex contraction. Thus, by locally modulating actomyosin-cortex-dependent surface tension and globally regulating osmotic pressure, cells can control their volume, shape and mechanical properties.

To analyse cell shape during mitosis, we simultaneously used atomic force microscopy (AFM), to measure cell height, and transmitted light microscopy, to measure cell width (Methods and Supplementary Fig. 1). Because we can determine the position of the cantilever with nanometre precision, this provides a similarly precise measure of the cell dimensions. Metaphase HeLa cells had a height-to-width ratio of 0.86 ± 0.04 (mean \pm s.d.; Supplementary Fig. 1b). Mitotic cells without retraction fibres were almost spherical, as were interphase cells detached with trypsin (Supplementary Fig. 1b, c). Therefore, we conclude that a detached, isolated cell will be nearly spherical, independent of its cell cycle phase. This suggests that loss of adhesion as cells enter mitosis permits cell rounding³.

A role for actin-based processes has previously been demonstrated in mitotic cell rounding^{1,4,6,8,9}. Therefore, we tested the role of the actin cytoskeleton in maintaining a spherical shape by adding cytochalasin D to rounded cells (Supplementary Fig. 1a, e). After treatment, both detached mitotic and interphase cells remained round. However, if retraction fibres were present, rounded cells sagged to height-to-width ratios of <0.5 on cytochalasin D treatment. Therefore, the actomyosin cytoskeleton is necessary for generating a rounding force against adhesion.

To quantify the force of cell rounding, a tipless cantilever was positioned over a prophase HeLa cell, 8 μm above the substrate (Fig. 1a), and held there while the cell underwent mitosis. We refer to this method as a 'constant-height assay'. When becoming rounder in prometaphase, the mitotic cell came in contact with the cantilever and the upward force that it exerted on the cantilever was measured with subnanonewton accuracy. Simultaneously, the cell's progression through mitosis was monitored using light microscopy (Fig. 1b). Within ~ 10 min after nuclear envelope breakdown, cells were cylindrical, and remained so until division. As cells progressed through

prometaphase and into metaphase, the force exerted on the cantilever increased. Because cortical tension was uniform across the cell until anaphase¹⁰ (Supplementary Fig. 2), we were able to normalize force by

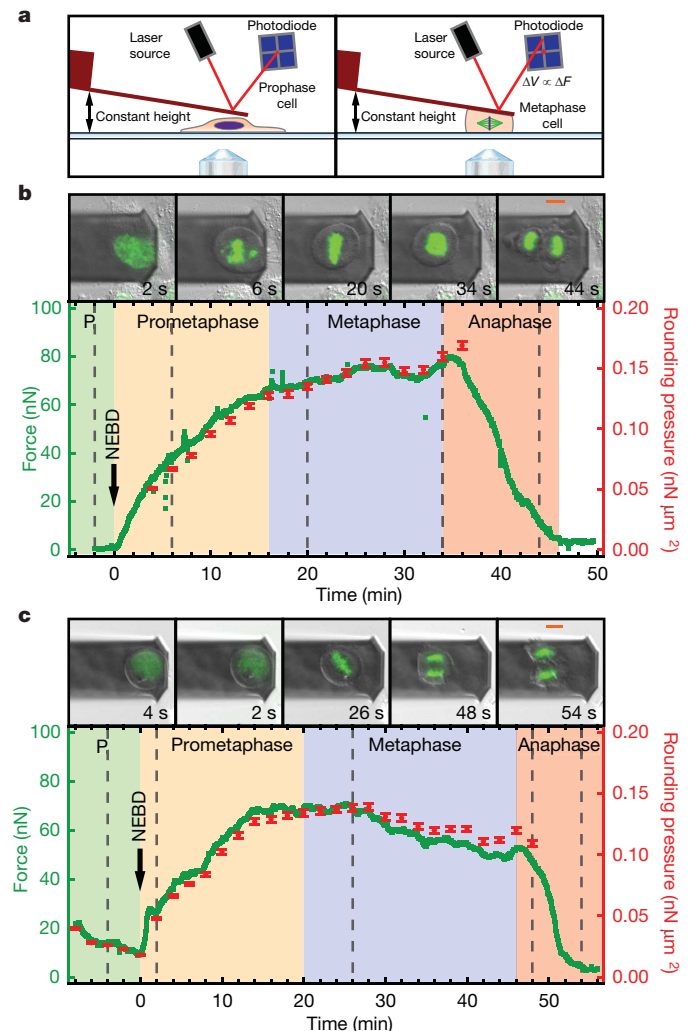


Figure 1 | Cells exert an increased rounding pressure in mitosis.

a, Constant-height assay (Methods Summary). V , voltage signal at the AFM photodiode; F , force. **b**, Overlaid differential interference contrast (DIC) and histone H2B/green fluorescent protein (GFP) images of a mitotic HeLa cell at the times indicated by the grey dashed lines. Graphed is the measured upward force (green) and calculated rounding pressure (red), which could be derived only while the cell was cylindrical (Methods). Time zero denotes nuclear envelope breakdown (NEBD). Mitotic phases are as follows: prophase (P, green), prometaphase (orange), metaphase (blue) and anaphase (red). **c**, As in **b** but for a mitotic cell pre-rounded with trypsin treatment before NEBD. Error bars, $\pm 2\%$ (based on measurement uncertainty from DIC images); scale bars, 10 μm .

¹ETH Zürich, Department of Biosystems Science and Engineering, CH-4058 Basel, Switzerland. ²Biotechnology Center, University of Technology Dresden, D-01307 Dresden, Germany. ³Max-Planck-Institute of Molecular Cell Biology and Genetics, D-1307 Dresden, Germany.

dividing it by the cross-sectional area of the cell, providing a cell 'rounding pressure'. In metaphase, the rounding pressure reached a maximum at $0.14 \pm 0.04 \text{ nN } \mu\text{m}^{-2}$ ($n = 83$). At anaphase, before ingression of the cleavage furrow, there was a transient rise in the rounding pressure. Concomitant with cytokinesis, force decreased, and the daughter cells flattened and finally lost contact with the cantilever. We also showed that individual G2/prophase cells that had been pre-rounded with trypsin and progressed into mitosis (Fig. 1c) increased their rounding pressure by more than 3-fold, despite cell size remaining relatively constant (Supplementary Fig. 3). Thus, as cells enter mitosis, they can exert considerable forces against external objects. These changes in mechanical properties resemble those of early studies on sea urchin eggs¹¹.

To determine what mechanisms generate the increased rounding force during mitosis, we examined the forces exerted by mitotic cells in the presence of various perturbants of the cytoskeleton. All actomyosin inhibitors tested significantly reduced the rounding pressure of mitotic cells (Fig. 2a). In contrast, perturbing microtubule dynamics increased the exerted pressure (Fig. 2a), perhaps because Rho/Rac guanine nucleotide exchange factor 2 was no longer inhibited¹². Therefore, an intact actomyosin cortex, but not the microtubule cytoskeleton, is required for rounding cells to generate a rounding pressure against an external impediment.

While performing the constant-height assay on mitotic cells using intermediate concentrations of latrunculin A (40–100 nM), we noticed oscillations in the rounding force, which correlated with blebbing (Fig. 2b). The rounding pressure decreased while blebs expanded, and recovered during bleb retraction (Fig. 2b). A bleb forms when a section of membrane detaches from the actomyosin cortex¹³ and retracts when the actomyosin cortex reassembles underneath the membrane and pulls the bleb back into the main cell body¹⁴. The concurrence of bleb formation and the decrease in rounding pressure suggests that the cell was under hydrostatic pressure: the pressure inside the cells pushing the cantilever upward was partially released when a bleb formed¹⁵. This interpretation is supported by recent measurements quantitatively relating cortical actomyosin tension with bleb formation¹⁶.

The force on the cantilever could be a result of osmotic pressure. If the osmolarity is higher inside the cell than outside, water will flow into the cell and generate a hydrostatic pressure. To test this idea, we modulated the osmolarity of the medium. Introduction of hypotonic medium ($-\Delta 100 \text{ mosM l}^{-1}$) led to an immediate increase in the volume of metaphase cells ($40 \pm 6\%$; $n = 9$), indicating that water entered the cells (Fig. 3a). This was accompanied by a concurrent increase in the measured rounding pressure ($76 \pm 20\%$; $n = 9$), presumably because the intracellular pressure increased. Within 3 min of the osmolarity changing, the cell volume and rounding pressure returned to close to their original values. This is probably because, in response to increased osmotic pressure, regulatory volume decrease causes cells to release ions¹⁷. Conversely, when hypertonic medium ($+\Delta 200 \text{ mosM l}^{-1}$) was introduced (Fig. 3b) the changes in volume ($-24 \pm 4\%$; $n = 9$) and rounding pressure ($-30 \pm 14\%$; $n = 9$) were in the opposite direction. Again, the cells recovered the original rounding pressure and volume, presumably because regulatory volume increase triggers the influx of osmolytes¹⁷.

Because ion transporters at the plasma membrane increase intracellular osmotic pressure and restore the volume of cells immediately after hypertonic challenge, we reasoned that they might also contribute to the increased rounding pressure seen in mitosis (Fig. 1). Therefore, we tested the effect of inhibiting ion transporters important in regulatory volume increase¹⁷. Among the inhibitors tested, an inhibitor of Na^+/H^+ antiporters, ethylisopropylamiloride, caused the greatest decrease in rounding pressure ($-53 \pm 10\%$; $n = 19$) and volume ($-8 \pm 2\%$; $n = 19$) (Fig. 3c). The exchange of a proton with a Na^+ ion increases the intracellular osmolarity because pH is strongly buffered in the cytoplasm; thus, a Na^+ ion has a greater effect on osmolarity than a

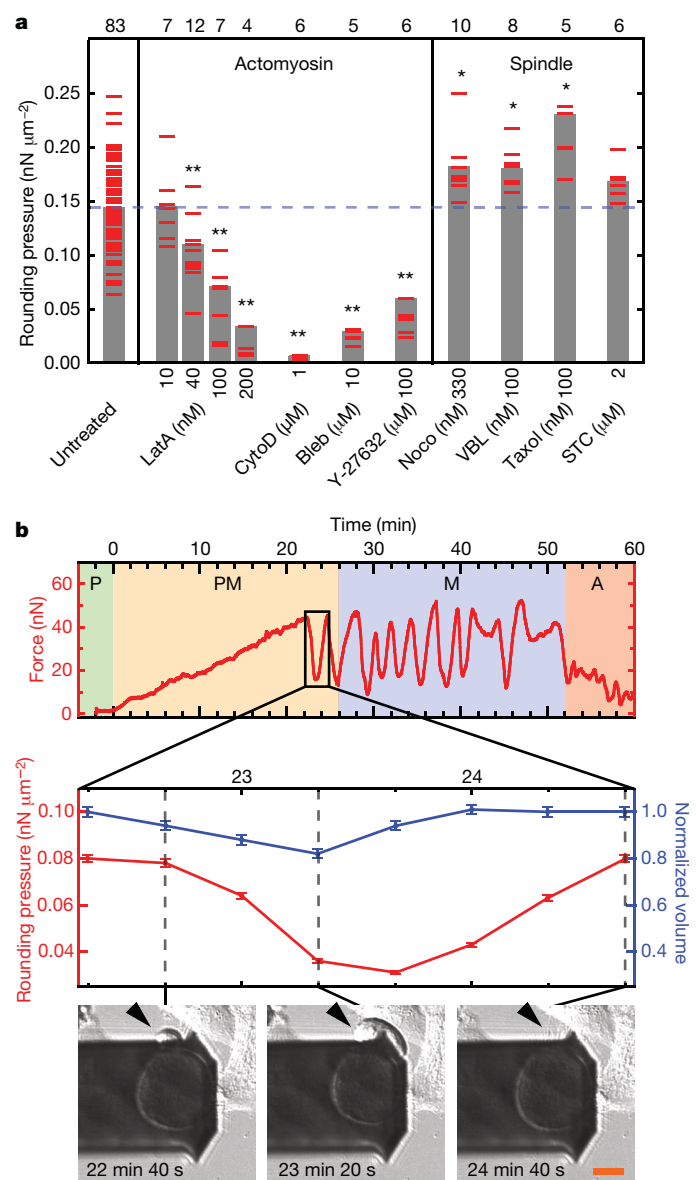


Figure 2 | Mitotic cells require a functional actin cytoskeleton to generate rounding pressure. **a**, Maximum rounding pressures generated by mitotic cells while incubated with inhibitors of the actomyosin system (latrunculin A (Lata, $n = 7, 12, 7$, and 4 , respectively), cytochalasin D (CytoD, $n = 6$), blebbistatin (Bleb, $n = 5$), and Y-27632 ($n = 6$) or microtubule spindle perturbants (nocodazole (Noco, $n = 10$), vinblastine (VBL, $n = 8$), Taxol (TXL, $n = 5$) and S-trityl-cysteine (STC, $n = 6$)). Red marks are maximum rounding pressures generated by single cells. Grey bars denote averages. n values are shown above each category. $*P < 0.05$, $**P < 0.001$. **b**, Top: rounding force exerted by a mitotic cell incubated with 40 nM Lata. Bottom: rounding pressure (red) and cell volume (not including the bleb, blue) during an oscillation, with corresponding DIC images showing bleb expansion and retraction. The drop in rounding pressure ($< -50\%$) cannot be accounted for by the volume decrease in the main cell body ($\sim -15\%$) indicating a decrease in intracellular pressure. A, anaphase; M, metaphase; PM, prometaphase. Error bars, $\pm 2\%$ (based on measurement uncertainty from DIC images); scale bar, $10 \mu\text{m}$.

proton^{17,18}. These results suggest the Na^+/H^+ antiporter increases mitotic rounding pressure, and agree well with data showing that the Na^+/H^+ antiporter SLC9A1 is activated at the G2–M transition¹⁹.

We next perturbed the ion gradients across the plasma membrane using *Staphylococcus aureus* α -toxin, which confers permeability to monovalent cations²⁰. When added to metaphase cells, it caused a decrease in both volume ($-42 \pm 4\%$; $n = 11$) and rounding pressure

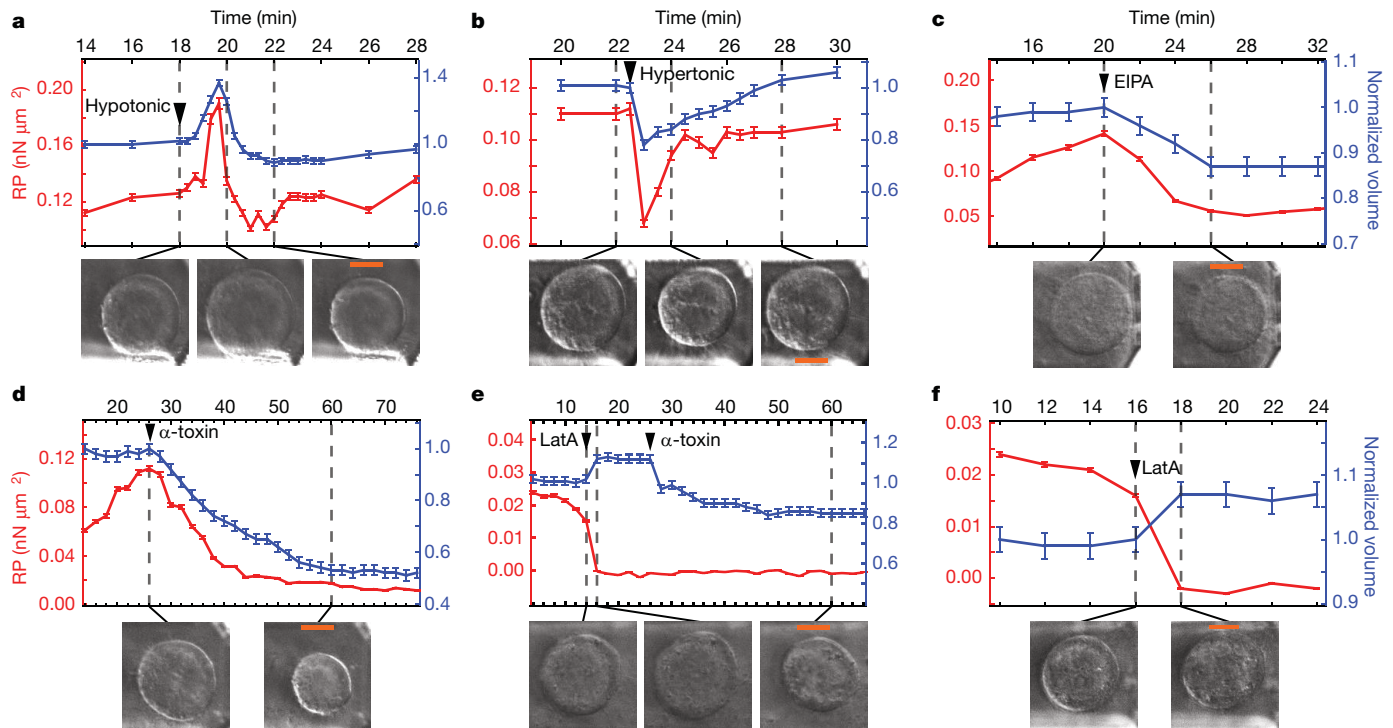


Figure 3 | The actomyosin cortex contracts against an intracellular osmotic pressure. Representative rounding pressure (RP) and cell volume time courses for mitotic cells subjected to the following perturbations: hypotonic ($-\Delta 100 \text{ mosM l}^{-1}$) medium ($n = 9$) (a); hypertonic ($+\Delta 200 \text{ mosM l}^{-1}$; +3% xylose) medium ($n = 9$) (b); 50 μM ethylisopropylamiloride (EIPA, $n = 19$)

(c); 60 $\mu\text{g ml}^{-1}$ α -toxin ($n = 11$) (d); 1 μM latrunculin A (14 min) then α -toxin (26 min, $n = 9$) (e); and 1 μM latrunculin A ($n = 36$) (f). In e and f, pre-treatment with 5 μM blebbistatin prevented latrunculin-A-induced blebbing. Time zero represents NEBD. Error bars, $\pm 2\%$ (based on measurement uncertainty from DIC images); scale bars, 10 μm .

($-88 \pm 8\%$; $n = 11$) (Fig. 3d). Another pore-forming toxin, *Escherichia coli* haemolysin A (HlyA), which also renders the plasma membrane permeable to cations²¹, also reduced the volume and rounding pressure of mitotic cells (Supplementary Fig. 4). Disruption of the actomyosin cortex with a combination of blebbistatin and latrunculin A before *S. aureus* α -toxin treatment reduced the toxin's ability to decrease cell volume (Fig. 3e), suggesting a contribution of the actomyosin cortex. We do not know whether this was due to an inward pressure caused by the actomyosin cortex or to more indirect mechanisms of volume regulation.

To further probe the link between osmotic pressure and actomyosin contraction, we performed experiments where we spontaneously abolished or stimulated the actomyosin cortex. If there is an intracellular pressure opposed by the actomyosin cortex, disruption of the cortex should result in dissipation of intracellular pressure and a small increase in cell volume. Indeed, when we treated mitotic cells first with blebbistatin, which inhibits myosin II contraction, and then with latrunculin A to depolymerize actin filaments, the volume of mitotic cells increased by $7 \pm 4\%$ ($n = 36$) and the mitotic rounding pressure was abolished (Fig. 3f). To study the converse case, we then looked at the effects of instantaneous activation of the actomyosin cortex. To do this we took advantage of blebbistatin's propensity to be inactivated by blue light²². When we photoinactivated blebbistatin, mitotic cells responded with an increase in rounding pressure and a decrease in volume (Supplementary Fig. 5). We conclude that stimulating contraction of the actomyosin cytoskeleton increases rounding pressure and decreases volume, whereas disrupting actomyosin activity reduces rounding pressure and increases volume.

Our experiments show that perturbation of osmotic gradients, associated transporters and the actomyosin cortex caused changes in both volume and rounding pressure (Fig. 4). When osmotic pressure was reduced, rounding pressure and volume decreased (Fig. 4a,

b, lower left quadrants). Conversely, if osmotic pressure was increased, rounding pressure and volume increased (Fig. 4a, b, upper right quadrants). When the actomyosin cytoskeleton was abolished, cell volume increased while rounding pressure decreased (Fig. 4a, b, lower right quadrants). However, rounding pressure increased as volume decreased after actomyosin contraction was stimulated (Fig. 4a, b, upper left quadrants). In conclusion, these results reveal that the actomyosin cortex contracts against an opposing intracellular osmotic pressure.

We propose the following model for the active processes that drive cell rounding during mitosis. As cells enter mitosis, de-adhesion from the substrate allows cells to become rounder³ (Supplementary Fig. 1). At the same time, cells increase their intracellular pressure (Fig. 1), presumably to drive rounding in a tightly packed tissue environment. Our model implies that intracellular pressure is at least equal to the measured rounding pressure, $\sim 150 \text{ Pa}$ ($0.15 \text{ nN } \mu\text{m}^{-2}$; Fig. 1). A pressure difference across the cell membrane of 100–500 Pa is thought to be sufficient to cause cell blebbing¹⁵ and is within the range typically measured in micropipette aspiration techniques²³ (1–1,000 Pa). A corollary to this model is that a non-homogeneous cortex results in dissimilar cell surface curvatures such as those observed in blebbing cells. In conjunction with an intracellular hydrostatic pressure, local modulation of cortical tension would allow cells to alter their shape^{24,25}, control their motion^{25–29} and govern the mechanics of mitosis¹⁵.

Pressure gradients are known to drive shape changes in organisms with cell walls³⁰. Our experiments support the idea that the actomyosin cortex behaves like an internal cell wall that directs osmotic expansion to control animal cell shape^{25–29}. Given the intricate shapes microorganisms and plants are able to achieve using turgor pressure, it is perhaps not surprising that animal cells have also evolved a mechanism that makes use of osmotic pressure.

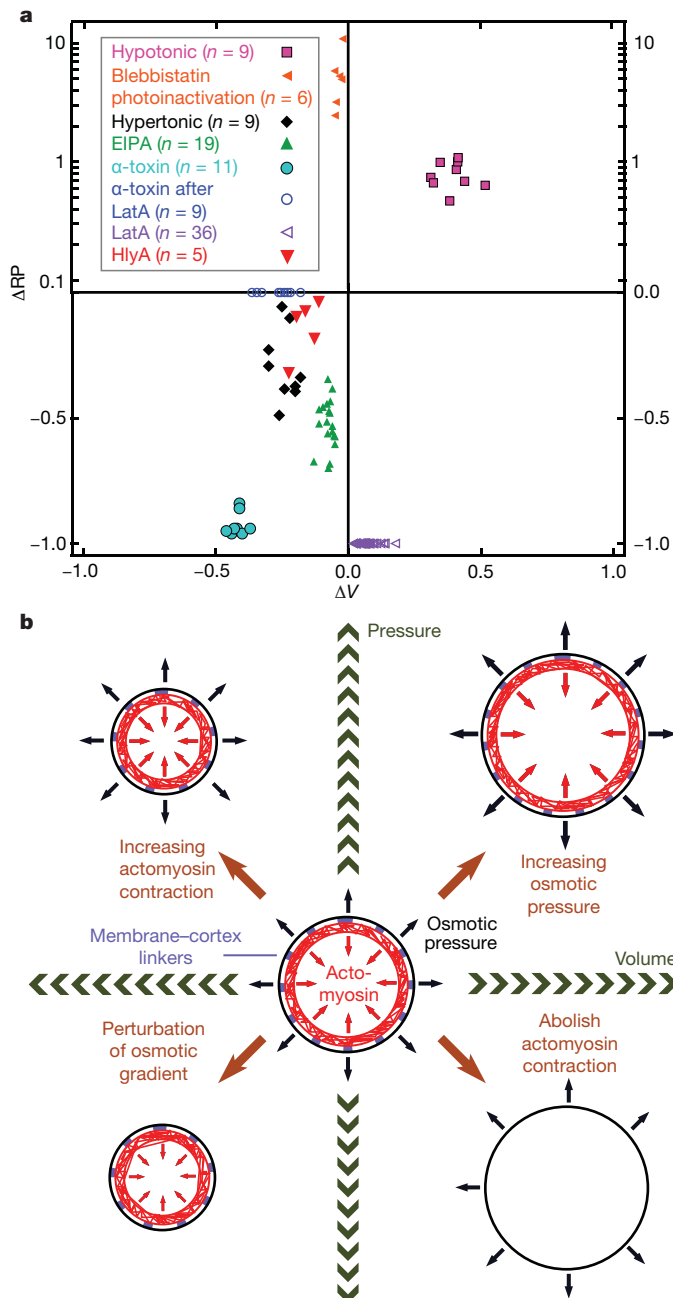


Figure 4 | Animal cells control shape in mitosis by modulating intracellular pressure in conjunction with actomyosin activity. **a**, Relative maximum changes in rounding pressure (ΔRP) and normalized volume (ΔV) of mitotic cells upon treatment with the indicated perturbations. Concentrations used are those indicated in Fig. 3 legend, $2 \mu\text{g ml}^{-1}$ for haemolysin A (HlyA) and $10 \mu\text{M}$ for blebbistatin photoinactivation, which involved a 1-s exposure to blue light. ΔRP is shown on a logarithmic scale above $\Delta RP = 0.1$. n values are displayed in the key. **b**, Uniform actomyosin contractile tension (red) is balanced by an outward-directed, intracellular osmotic pressure (black). Membrane-cortex linkers (purple) couple these two elements. The higher the tension and pressure, the greater the cortex rigidity. Imbalances between tension and osmotic pressure cause changes in cell volume and rounding force.

METHODS SUMMARY

Cell culture. HeLa-Kyoto cells expressing a fluorescent histone construct (H2B-GFP) were grown to ~50% confluency on 24-mm-diameter glass coverslips for cantilever experiments. We used DMEM containing 4 mM sodium bicarbonate (PN:31600-083, Invitrogen) buffered with 20 mM HEPES for experiments.

Instrumentation. The experimental set-up consisted of an AFM (Nanowizard I, JPK Instruments) mounted on a Zeiss Axiovert 200M optical microscope. Tipless

cantilevers were ~250 μm long, 35 μm wide, 2 μm thick and made of pure silicon (NSC12-D/tipless/noAl, $k \approx 0.3 \text{ N m}^{-1}$, Mikromasch). A BioCell (JPK Instruments) maintained cells at 37 °C.

AFM constant-height assay. We positioned the end of the cantilever over a candidate cell, 8 μm above the substrate. A laser beam was used to monitor the position of the cantilever, which was calibrated and used to record forces generated by cells. Rounding pressure was derived by dividing the measured force by the horizontal cross-sectional area of the cell, which was measured using DIC images. Cell volume was determined by multiplying the cross-sectional area of the near-cylindrical cell by its height under the cantilever. The $\pm 2\%$ error in rounding pressure and volume shown in the figures is based on measurement uncertainty from DIC images.

In Fig. 2a, cells were pre-incubated with perturbants. In Fig. 3a, b, tonic shock was induced by exchanging the full volume of the AFM BioCell several times with WPI Aladdin push-pull pumps. In Fig. 3c–f, perturbants were added to the AFM BioCell with a microsyringe.

We determined the following mitotic phases from H2B-GFP images: prophase, condensed chromosomes but intact nucleus; prometaphase, nuclear envelope breakdown; metaphase, chromosomes aligned to form a metaphase plate; anaphase, two sets of chromosomes separated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 November 2009; accepted 1 November 2010.

Published online 2 January 2011.

- Cramer, L. P. & Mitchison, T. J. Investigation of the mechanism of retraction of the cell margin and rearward flow of nodules during mitotic cell rounding. *Mol. Biol. Cell* **8**, 109–119 (1997).
- Gibson, M. C., Patel, A. B., Nagpal, R. & Perrimon, N. The emergence of geometric order in proliferating metazoan epithelia. *Nature* **442**, 1038–1041 (2006).
- Harris, A. Location of cellular adhesions to solid substrata. *Dev. Biol.* **35**, 97–114 (1973).
- Carreno, S. *et al.* Moesin and its activating kinase Slik are required for cortical stability and microtubule organization in mitotic cells. *J. Cell Biol.* **180**, 739–746 (2008).
- Kunda, P. & Baum, B. The actin cytoskeleton in spindle assembly and positioning. *Trends Cell Biol.* **19**, 174–179 (2009).
- Kunda, P., Pelling, A. E., Liu, T. & Baum, B. Moesin controls cortical rigidity, cell rounding, and spindle morphogenesis during mitosis. *Curr. Biol.* **18**, 91–101 (2008).
- Théry, M. & Bornens, M. Cell shape and cell division. *Curr. Opin. Cell Biol.* **18**, 648–657 (2006).
- Fujibuchi, T. *et al.* AIP1/WDR1 supports mitotic cell rounding. *Biochem. Biophys. Res. Commun.* **327**, 268–275 (2005).
- Maddox, A. S. & Burridge, K. RhoA is required for cortical retraction and rigidity during mitotic cell rounding. *J. Cell Biol.* **160**, 255–265 (2003).
- Matzke, R., Jacobson, K. & Radmacher, M. Direct, high-resolution measurement of furrow stiffening during division of adherent cells. *Nature Cell Biol.* **3**, 607–610 (2001).
- Hiramoto, Y. Mechanical properties of sea urchin eggs. I. Surface force and elastic modulus of the cell membrane. *Exp. Cell Res.* **32**, 59–75 (1963).
- Krendel, M., Zenke, F. T. & Bokoch, G. M. Nucleotide exchange factor GEF-H1 mediates cross-talk between microtubules and the actin cytoskeleton. *Nature Cell Biol.* **4**, 294–301 (2002).
- Cunningham, C. C. Actin polymerization and intracellular solvent flow in cell surface blebbing. *J. Cell Biol.* **129**, 1589–1599 (1995).
- Charras, G. T., Hu, C. K., Coughlin, M. & Mitchison, T. J. Reassembly of contractile actin cortex in cell blebs. *J. Cell Biol.* **175**, 477–490 (2006).
- Charras, G. T., Coughlin, M., Mitchison, T. J. & Mahadevan, L. Life and times of a cellular bleb. *Biophys. J.* **94**, 1836–1853 (2008).
- Tinevez, J. Y. *et al.* Role of cortical tension in bleb growth. *Proc. Natl Acad. Sci. USA* **106**, 18581–18586 (2009).
- Lang, F. *Mechanisms and Significance of Cell Volume Regulation* (Karger, 2006).
- Wehner, F., Olsen, H., Tinel, H., Kinne-Saffran, E. & Kinne, R. K. Cell volume regulation: osmolytes, osmolyte transport, and signal transduction. *Rev. Physiol. Biochem. Pharmacol.* **148**, 1–80 (2003).
- Putney, L. K. & Barber, D. L. Na-H exchange-dependent increase in intracellular pH times G2/M entry and transition. *J. Biol. Chem.* **278**, 44645–44649 (2003).
- Valeva, A. *et al.* Staphylococcal alpha-toxin: repair of a calcium-impermeable pore in the target cell membrane. *Mol. Microbiol.* **36**, 467–476 (2000).
- Koschinski, A. *et al.* Why *Escherichia coli* alpha-hemolysin induces calcium oscillations in mammalian cells—the pore is on its own. *FASEB J.* **20**, 973–975 (2006).
- Kolega, J. Phototoxicity and photoinactivation of blebbistatin in UV and visible light. *Biochem. Biophys. Res. Commun.* **320**, 1020–1025 (2004).
- Hochmuth, R. M. Micropipette aspiration of living cells. *J. Biomech.* **33**, 15–22 (2000).
- Salbreux, G., Joanny, J. F., Prost, J. & Pullarkat, P. Shape oscillations of non-adhering fibroblast cells. *Phys. Biol.* **4**, 268–284 (2007).

25. Blaser, H. *et al.* Migration of zebrafish primordial germ cells: a role for myosin contraction and cytoplasmic flow. *Dev. Cell* **11**, 613–627 (2006).
26. Bereiter-Hahn, J. Mechanics of crawling cells. *Med. Eng. Phys.* **27**, 743–753 (2005).
27. Charras, G. & Paluch, E. Blebs lead the way: how to migrate without lamellipodia. *Nature Rev. Mol. Cell Biol.* **9**, 730–736 (2008).
28. Mitchison, T. J., Charras, G. T. & Mahadevan, L. Implications of a poroelastic cytoplasm for the dynamics of animal cell shape. *Semin. Cell. Dev. Biol.* **19**, 215–223 (2008).
29. Keren, K., Yam, P. T., Kinkhabwala, A., Mogilner, A. & Theriot, J. A. Intracellular fluid flow in rapidly moving cells. *Nature Cell Biol.* **11**, 1219–1224 (2009).
30. Harold, F. M. To shape a cell: an inquiry into the causes of morphogenesis of microorganisms. *Microbiol. Rev.* **54**, 381–431 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements DFG, BMBF and SNF supported this project. JSPS supported Y.T. A.A.H. is funded by the Max Planck Society. We thank S. Bhakdi for toxins and advice on

their use, T. J. Mitchison for extensive discussions on osmotic pressure and critical reading of the manuscript, M. Krieg for valuable insights into cell blebbing, and B. Baum, C. Brangwynne, S. Grill, A. Helenius, J. Howard, F. Jülicher, Z. Maliga and E. Paluch for discussions and critical reading of the manuscript.

Author Contributions M.P.S., D.J.M., Y.T. and A.A.H. designed and implemented the assay. M.P.S. performed the experiments except for the Young's modulus measurements, which were made by J.H. S.P.R. contributed to Fig. 2 and Supplementary Fig. 6. Y.T. produced cell lines. J.H. designed the toxin experiments. M.P.S. analysed data and created the figures. M.P.S., J.H., D.J.M. and A.A.H. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.A.H. (hyman@mpi-cbg) or D.J.M. (daniel.muller@bsse.ethz.ch).

METHODS

Cell culture. HeLa-Kyoto cells expressing a histone H2B-eGFP construct (H2B-GFP) were used³¹. In some experiments, cells additionally expressing mCherry-CAAX, fluorescently marking the plasma membrane, were used. To construct the mCherry-CAAX plasmid (TH0477), a DNA sequence (5'-TGCATGAGCTGC AAGTGTGTGCTGTCC-3') from the carboxy terminus of the rat c-H-Ras 1 gene was inserted to the C terminus of the mCherry gene³². The mCherry-CAAX plasmid was transfected into H2B-GFP cells and selected by puromycin resistance. Cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS), 2 mM GlutaMAX, 100 units ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin and 0.5 mg ml⁻¹ Geneticin (all Invitrogen) at 37 °C in a 5% CO₂ environment. DMEM used for AFM experiments (cat. no. 31600-083, Invitrogen) contained only 4 mM sodium bicarbonate and was instead buffered with 20 mM HEPES/NaOH to pH 7.2. When using dynasore and trypsin, NuSerum (Fischer Scientific) was substituted for FBS at the corresponding concentration. For AFM experiments, cells were plated on 24-mm-diameter glass coverslips (Marienfeld) and grown for 2 days until they reached ~50% confluency.

Perturbants. Chemical inhibitors were acquired from Sigma-Aldrich except jasplakinolide (Merck) and trypsin (Invitrogen) and used at the indicated concentrations. In cases where perturbants were added during a measurement, a microsyringe (Hamilton) was used to inject agents into the AFM BioCell or the full volume was exchanged several times with a WPI Aladdin push-pull pump set-up (Supplementary Fig. 7a). Supplementary Table 1 lists all perturbants used and how they were added to cells, and provides a short description of what the perturbant does.

Pore-forming-toxin experiments. *S. aureus* α -toxin stock was made at 2 mg ml⁻¹ in PBS. Owing to having limited amounts of toxin stock, we added 100 µl of a $\times 5$ concentrate to 400 µl of medium to arrive at the final concentration. Concentrate was added with a tube and syringe in the manner shown in Supplementary Fig. 7a. At a given concentration of α -toxin, more crowded coverslips seemed to dilute the potency of the toxin. Thus, for toxin experiments we attempted to keep the cell confluency constant at 40–50%. Concentrations at and below 20 µg ml⁻¹ had no noticeable effect on cells. We found a dosage of 40–80 µg ml⁻¹ caused effects such as those seen in Fig. 3d.

Haemolysin A stock was made up at 910 µg ml⁻¹ in 8 M guanine/HCl solution. Again, owing to having limited amounts of toxin stock we used the $\times 5$ concentrate 'addition' method described for α -toxin. We found that a final concentration of 1 µg ml⁻¹ had little effect on cells. However, using 2–4 µg ml⁻¹ we obtained typical results seen in Supplementary Fig. 4.

Osmolarity change experiments. For osmolarity change experiments, the volume of the AFM BioCell (~400 µl) was exchanged four times at a flow rate of 2,500 µl min⁻¹ using a WPI Aladdin push-pull pump. For hypotonic exchanges, DMEM (290 mosM kg⁻¹) was diluted with water to 190 mosM kg⁻¹. In the case of hypertonic exchanges, DMEM was supplemented with xylose to increase osmolarity to 490 mosM kg⁻¹. To minimize disturbance of the AFM laser, the refractive index of all solutions was matched with Ficoll (Sigma-Aldrich), an inert sucrose polymer with negligible effect on osmolarity in solution. **Mid-experiment introduction of perturbants: balancing refractive index and osmolarity difference.** In cases where the perturbant mixture differed from the original medium in refractive index, we added Ficoll to the solution having lower refractive index until the solutions matched. 1% w/v Ficoll increased refractive index by 0.0013 with negligible change to osmolarity.

In cases where the perturbant mixture and original medium had an unintended mismatch in osmolarity, we used xylose to raise the osmolarity of the lower-osmolarity solution. 1% w/v xylose increases osmolarity by 70 mosM l⁻¹ but also increased refractive index by 0.0013. In such cases, Ficoll was additionally used to match refractive index. In all cases, Ficoll and xylose never exceeded 3% w/v and did not negatively effect progression through mitosis or the rounding forces measured.

Instrumentation. The experimental set-up consisted of an AFM (Nanowizard I, JPK Instruments) mounted on an Axiovert 200M optical microscope (Carl Zeiss). A BioCell (JPK Instruments) allowed cells to be cultured at 37 °C during experiments. Tipless cantilevers were ~250 µm long, 35 µm wide, 2 µm thick and made of pure silicon (NSC12-D/tipless/noAl from Mikromasch) and had a nominal force constant of 0.3 N m⁻¹. Cantilevers were calibrated using the thermal noise method³³.

AFM constant-height assay. The procedure for an 8-µm constant-height assay of a mitotic HeLa cell was as follows. The height of the substrate adjacent to a cell was determined. Then the cantilever end was positioned over the cell, 8 µm above the substrate. Prophase cells typically had heights of <7 µm, and were therefore initially not in contact with the cantilever. The force exerted by the rounding mitotic cell was sensed by the cantilever and recorded over time. Rounding pressure was derived by dividing the force measured by the horizontal cross-sectional

area of the cell, which was measured from DIC images of the cell (Supplementary Fig. 7c). Cell volume was determined by multiplying the cross-sectional area of the near-cylindrical cell by its height under the cantilever (Supplementary Fig. 7c). The $\pm 2\%$ error on rounding pressure and volume is based on measurement uncertainty from DIC images. After a constant-height assay, the height of the substrate was re-measured to assess mechanical drift. Experiments with more than 10% drift in relative height were discarded. Cantilever deflection drift was usually <2 nN per hour and could be neglected when considering the magnitude of forces measured.

Cell height and shape measurements. Cell heights were determined using AFM. Because the cell is soft, particularly when treated with actin perturbants, the cell height was determined by extracting a contact point from force-distance curves. As the cantilever approached the cell, a small force, of ~100–200 pN, is registered at the cantilever upon contact. The vertical distance between this contact point and the substrate was defined as the cell height. To obtain a cell height-to-width ratio, the cell's width was measured using DIC microscopy images.

Optical microscopy. For optical microscopy, a Plan Apochromat $\times 20/0.8$ objective lens (Zeiss) was used. Images were acquired with a CoolSNAP cf camera (Roper Scientific) driven by METAMORPH software (Visitron). An alternative set-up featuring an MRM camera (Zeiss) and AXIOVISION software (Zeiss) was used to acquire and analyse images in some experiments, including the ApoTome optical sectioning shown in Supplementary Fig. 7b. For H2B-GFP and mCherry-CAAX imaging, FITC and Texas red filter sets were respectively used.

Mitotic phase assignment. H2B-GFP images were used to determine the phase of mitosis according to the following criteria: prophase cells contain condensed chromosomes surrounded by a nuclear envelope; prometaphase starts with nuclear envelope breakdown; metaphase is when chromosomes align at the metaphase plate; and anaphase is when the two sets of chromosomes separate.

Cortical stiffness measurements. The measurement was made by indenting the cell cortex using a 5-µm-diameter bead glued to a NPO tipless cantilever (Veeco; nominal stiffness, 50 mN m⁻¹) with a force of 2 nN. The Young's modulus at each indentation was calculated by least-squares fitting a parabolic curve to the force-indentation curve at contact³⁴. The method was adapted from ref. 35. The bead diameter was determined using its DIC image and the cantilever spring constant was determined as described above. We assumed the cell to have a Poisson's ratio of 0.5 (ref. 35).

Constant-height assay design. We started these experiments by using soft cantilevers (Nanoworld TL1, $k \approx 50$ mN m⁻¹). However they could neither measure nor apply forces over 20 nN; therefore, we changed to stiff cantilevers. We prefer them because they allowed us to measure a wide range of forces, from tens of piconewtons to hundreds of nanonewtons. Using such cantilevers in an 8-µm constant-height assay (Supplementary Fig. 7) allowed us to sample force, rounding pressure and volume continuously. This was key to understanding perturbations where volume and pressure respond quickly (Figs 2 and 3). Additionally, we compared the effect of long-term constraint under the cantilever with measurements on a shorter timescale using a simple rheological compression assay (Supplementary Fig. 8c). We plotted our two data sets, constant-height assays ($n = 123$) and compression assays ($n = 100$), as force versus deformation (Supplementary Fig. 8a) and contact stress versus strain (Supplementary Fig. 8b). Because we get similar results whether we 'constrain the cell' or make a simple mechanical measurement on an already round one, we conclude that our constant-height assay is applicable to studying mitotic cell rounding over a broad range of deformations and timescales.

Cell viability under the cantilever. To investigate the effect of constraining cell shape in our assay on progression through mitosis, we measured the times taken for individual cells to progress from prophase to anaphase (Supplementary Fig. 9a). Cells either subjected to a constant force of 1–50 nN or constrained to a constant height of 8 µm progressed through mitosis unaffected. However, addition of cytochalasin D to cells in an 8-µm constant-height assay slightly extended the time between nuclear envelope breakdown and anaphase. Additionally, the height-to-width ratio attained by mitotic cells becoming rounder against constant forces is plotted in Supplementary Fig. 9b.

- Neumann, B. *et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. *Nature Methods* **3**, 385–390 (2006).
- Ruta, M. *et al.* Nucleotide sequence of the two rat cellular rasH genes. *Mol. Cell. Biol.* **6**, 1706–1710 (1986).
- Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64**, 1868–1873 (1993).
- Hertz, H. Über den Kontakt elastischer Körper. *J. Reine Angew. Math.* **92**, 156–172 (1881).
- Radmacher, M., Fritz, M. & Hansma, P. K. Imaging soft samples with the atomic force microscope: gelatin in water and propanol. *Biophys. J.* **69**, 264–270 (1995).

c-Jun N-terminal phosphorylation antagonises recruitment of the Mbd3/NuRD repressor complex

Cristina Aguilera^{1*}, Kentaro Nakagawa^{1,2*}, Rocio Sancho¹, Atanu Chakraborty¹, Brian Hendrich³ & Axel Behrens¹

AP-1 (activator protein 1) activity is strongly induced in response to numerous signals, including growth factors, cytokines and extracellular stresses¹. The proto-oncoprotein c-Jun belongs to the AP-1 group of transcription factors and it is a crucial regulator of intestinal progenitor proliferation and tumorigenesis²⁻⁴. An important mechanism of AP-1 stimulation is phosphorylation of c-Jun by the Jun amino-terminal kinases (JNKs)¹. N-terminal phosphorylation of the c-Jun transactivation domain increases target gene transcription^{5,6}, but a molecular explanation was elusive. Here we show that unphosphorylated, but not N-terminally phosphorylated c-Jun, interacts with Mbd3 and thereby recruits the nucleosome remodelling and histone deacetylation (NuRD) repressor complex. Mbd3 depletion in colon cancer cells increased histone acetylation at AP-1-dependent promoters, which resulted in increased target gene expression. The intestinal stem cell marker *Igr5* was identified as a novel target gene controlled by c-Jun/Mbd3. Gut-specific conditional deletion of *mbd3* (*mbd3*^{AG/AG} mice) stimulated c-Jun activity and increased progenitor cell proliferation. In response to inflammation, *mbd3* deficiency resulted in colonic hyperproliferation and *mbd3*^{AG/AG} mice showed markedly increased susceptibility to colitis-induced tumorigenesis. Notably, concomitant inactivation of a single allele of *c-jun* reverted physiological and pathological hyperproliferation, as well as the increased tumorigenesis in *mbd3*^{AG/AG} mice. Thus the transactivation domain of c-Jun recruits Mbd3/NuRD to AP-1 target genes to mediate gene repression, and this repression is relieved by JNK-mediated c-Jun N-terminal phosphorylation.

We described a yeast three-hybrid approach to identify proteins that interact with c-Jun depending on its phosphorylation status previously⁷. Mbd3 was cloned as an interactor that specifically bound to unphosphorylated c-Jun. Mbd3, together with Mbd1, 2 and 4, was originally characterized as a protein containing a region with high homology to the methyl-CpG-binding domain (MBD) of MeCP2. Mbd2 and Mbd3 assemble into mutually exclusive distinct NuRD complexes⁸. NuRD mediates gene repression through histone deacetylation via HDAC1 and 2, and chromatin remodelling ATPase activities through its CHD3 (Mi2 α) and CHD4 (Mi2 β) subunits^{9–11}. Whereas Mbd2 recruits NuRD to methylated DNA, the MBD of Mbd3 fails to bind methylated DNA^{12,13}. Thus, how Mbd3 interacts with chromatin to regulate transcription was not known.

Overexpressed c-Jun co-immunoprecipitated with Mbd3 but JNK activation by anisomycin (Ans) reduced the interaction of wild-type c-Jun and Mbd3. In contrast, a c-Jun4A mutant, in which the four N-terminal residues phosphorylated by JNK, serines 63 and 73, and threonines 91 and 93, are mutated into alanines, remained bound to Mbd3 despite high JNK activity (Fig. 1a). c-Jun4A also interacted more efficiently with CHD3 and MTA3, two major components of the NuRD complex (Supplementary Fig. 1a). Interaction of endogenous Mbd3 with c-Jun could not be detected when JNK activity was stimulated

by ultraviolet light, but pharmacological JNK inhibition (JNKi), which resulted in a significant decrease in c-Jun N-terminal phosphorylation (Supplementary Fig. 2a), stimulated Mbd3 binding to c-Jun (Supplementary Fig. 1b). Moreover, recombinant glutathione-S-transferase

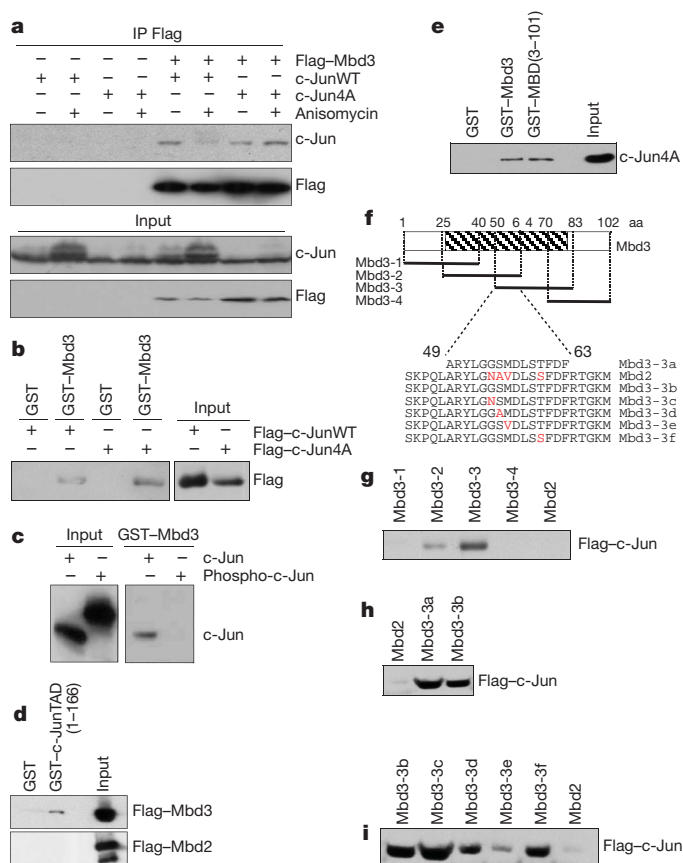


Figure 1 | Mbd3 interacts with unphosphorylated c-Jun through its MBD. **a**, HCT116 cells transfected with Myc-c-JunWT or Myc-c-Jun4A along with Flag-Mbd3 were treated with anisomycin for 1 h where indicated. Cell lysates were used to immunoprecipitate (IP) Flag-Mbd3. **b**, Pull-down experiment using GST-Mbd3 with HCT116 cell lysates transfected with c-JunWT or c-Jun4A. **c**, Pull-down experiment using GST-Mbd3 as a bait with recombinant c-Jun (c-Jun) or recombinant c-Jun that was phosphorylated *in vitro* by incubation with activated JNK1 protein (Phospho-c-Jun). **d**, Pull-down experiment using GST-JunTAD(1-166) with HCT116 cell lysates transfected with Flag-Mbd3 or Flag-Mbd2. **e**, Pull-down experiment using GST-Mbd3 or GST-MBD with HCT116 cell lysates transfected with Myc-c-Jun4A. **f**, Schematic overview of Mbd3 peptides covering the MBD, mutations are highlighted in red. **g-i**, Pull-down experiments using different biotinylated peptides shown in **f** with HEK293T cell lysates expressing Flag-c-Jun.

¹Mammalian Genetics Laboratory, Cancer Research UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. ²Department of Medical Biochemistry, Graduate School of Medicine, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan. ³Wellcome Trust Centre for Stem Cell Research, Medical Research Council Centre for Stem Cell Biology and Regenerative Medicine, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.

*These authors contributed equally to this work.

(GST)-Mbd3 interacted more efficiently with c-Jun4A than with wild-type c-Jun (Fig. 1b). Bacterially produced GST-Mbd3 bound to recombinant c-Jun, but this interaction was reduced by *in vitro* phosphorylation of recombinant c-Jun by activated JNK1 (Fig. 1c). GST pull-down assays showed that the transactivation domain (TAD, amino acid residues 1–166) of c-Jun was sufficient to bind Mbd3, but did not interact with the closely related Mbd2 protein (Fig. 1d). The delta (δ) domain of c-Jun (residues 31–60), which is required for JNK binding, was dispensable for Mbd3 interaction, and further mapping identified the region between residues 60 and 120, which encompasses the JNK phosphoacceptor residues, as the Mbd3 binding site of c-Jun (Supplementary Fig. 1c–e). Thus, Mbd3 binds directly and preferentially to the unphosphorylated TAD of c-Jun, and this interaction is disrupted by phosphorylation.

To determine the domain of Mbd3 involved in binding to c-Jun, serial deletion mutants of Mbd3 were generated (Supplementary Fig. 3a). The deletion of residues 40 to 80 of Mbd3, which encode part of the MBD, reduced Mbd3 binding to c-Jun (Supplementary Fig. 3b). Recombinant Mbd3 MBD (GST-MBD(3–101)) bound to c-Jun as efficiently as full-length Mbd3 (GST-Mbd3) (Fig. 1e). Using overlapping biotinylated peptides covering the Mbd3 MBD we identified the region from residue 49 to residue 63 as crucial for Mbd3 binding to c-Jun (Fig. 1f, g and Supplementary Fig. 4a). A 15-residue peptide covering this region was sufficient to mediate c-Jun interaction (Fig. 1h). This region differs only in four residues to the same region in Mbd2 and a peptide with the corresponding Mbd2 sequence did not bind c-Jun (Fig. 1h). Single amino acid changes revealed a large contribution of methionine 56 and to a lesser extent serine 55 of Mbd3 in c-Jun binding (Fig. 1i and Supplementary Fig. 4b). Thus, the MBD of Mbd3 is necessary and sufficient for c-Jun binding, indicating that the MBD of Mbd3 has evolved from a methyl-CpG-binding domain into a protein–protein interaction module.

To investigate the role of Mbd3 in regulating c-Jun function, we knocked-down *mbd3* with a small hairpin RNA (shRNA) construct (Supplementary Fig. 2b) and transfected it along with a luciferase reporter construct driven by an artificial promoter consisting of multimerized AP-1 site (AP-1 luciferase) to assess c-Jun transcriptional activity. The knock-down of *mbd3* induced an increase in reporter gene activity (Fig. 2a). Similar results were obtained using the urokinase-type plasminogen activator promoter (uPA-luciferase) (Supplementary Fig. 5a), a well known AP-1 target gene¹⁴.

c-jun transcription is induced very quickly in response to JNK activation and c-Jun autoregulates its own promoter via two proximal AP-1 binding sites¹⁵. Chemical JNKi (Fig. 2b) or expression of *jnk1/2*-specific (also known as *MAPK8/9*) shRNAs, which resulted in substantial depletion of JNK protein (Supplementary Fig. 2c), decreased endogenous *c-jun* mRNA levels but *c-jun* transcriptional repression by JNK inhibition was impaired in the absence of Mbd3 (Fig. 2b and Supplementary Fig. 5b). Moreover, overexpression of Flag-Mbd3 decreased transcriptional activation of *c-jun* and, to a lesser extent, *cd44*, which is also regulated by c-Jun⁴ (Supplementary Fig. 5d, e). Thus, depletion of Mbd3 results in derepression of AP-1 activity.

HCT116 colon cancer cells have constitutively high levels of activated JNK and phosphorylated c-Jun (Supplementary Fig. 2a) and chromatin immunoprecipitation (ChIP) for Mbd3 revealed minimal binding to a region covering the AP-1 sites of the *c-jun* promoter. However, JNKi and *jnk1/2* depletion resulted in a rapid and efficient recruitment of Mbd3 to the *c-jun* promoter (Fig. 2c and Supplementary Fig. 6a) and the *cd44* promoter (Supplementary Fig. 6b, c). Thus, the JNK signalling pathway regulates Mbd3 recruitment to c-Jun-dependent promoters.

ChIP analysis demonstrated that HDAC1 was also recruited to the *c-jun* promoter after JNKi and knock-down of Mbd3 prevented this recruitment (Fig. 2d). Unphosphorylatable c-Jun4A protein recruited Mbd3 more efficiently to the *c-jun* promoter compared to wild-type c-Jun (c-JunWT) (Supplementary Fig. 6d). c-Jun4A also recruited

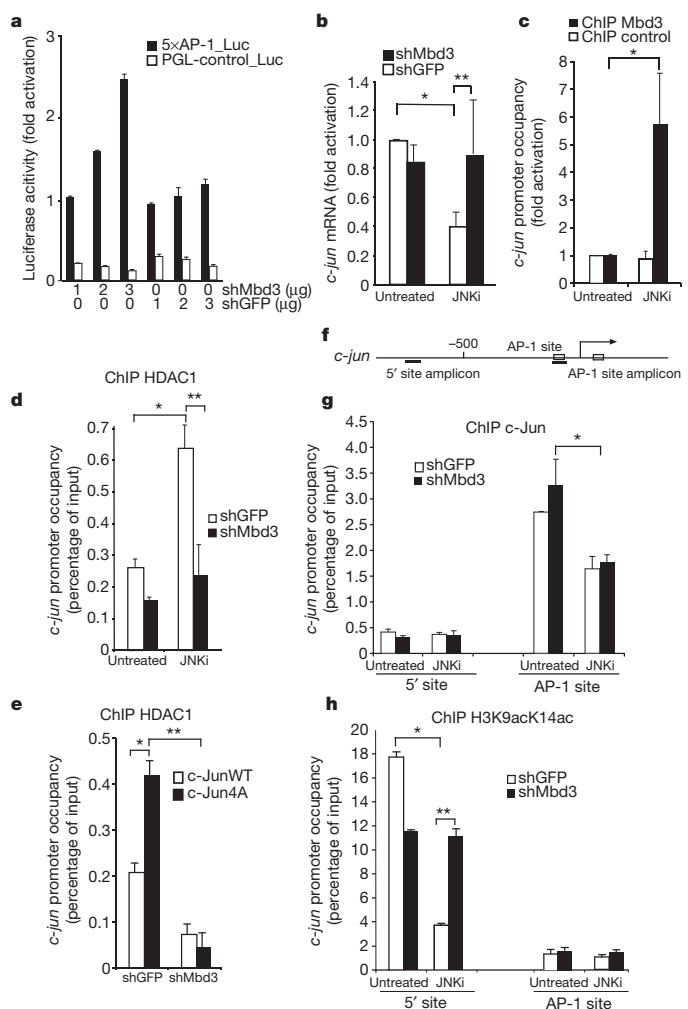


Figure 2 | Mbd3 represses *c-jun* transcription. **a**, Cells were transfected with a pSUPER vector expressing a shRNA specific for Mbd3 (shMbd3) or for GFP (shGFP). Activity of a 5×AP-1 luciferase reporter construct (5×AP-1_luc) or empty vector (PGL-control_Luc) relative to a thymidine kinase *Renilla* luciferase transfection control (tk-renilla) is shown ($n = 3$). **b**, mRNA levels of *c-jun* in the presence or absence of shMbd3 with JNKi or mock treatment ($n = 5$; $*P = 0.001$; $**P = 0.03$). **c**, ChIP for the AP-1 site in the *c-jun* promoter using Mbd3-specific or control antibody with or without JNKi ($n = 9$; $*P = 0.043$). **d**, HDAC1 ChIP for the AP-1 site in the *c-jun* promoter on cells expressing a shRNA specific for Mbd3 or for GFP ($*P = 0.004$; $**P = 0.0003$). **e**, HDAC1 ChIP for the *c-jun* AP-1 sites on cells expressing wild-type c-Jun (c-JunWT) or c-Jun4A ($*P = 0.02$; $**P = 0.01$). **f**, Schematic overview of the *c-jun* promoter. Underlined regions show the amplification region of each primer pair used. **g**, c-Jun ChIP for the 5' and AP-1 sites of the *c-jun* locus on cells expressing a shRNA specific for Mbd3 or for GFP ($*P = 0.02$). **h**, ChIP for acetylated histone 3 K9, 14 (H3K9acK14ac) as in **g** ($*P = 0.0002$; $**P = 0.006$). In all experiments error bars show s.e.m. and Student's *t*-test was used for statistical analysis. HCT116 cells were used throughout. JNKi treatment was for 2 h in **b**, and for 30 min in **c**, **d**, **g** and **h**.

more HDAC1 to the *c-jun* promoter, and this recruitment was prevented when Mbd3 was silenced (Fig. 2e).

Histone acetylation has an important role in *c-jun* transcriptional regulation¹⁶. It has been shown previously that the regulation of histone acetylation at the *c-jun* promoter is complex. Increased histone acetylation in response to JNK activation was observed in a more distal region of the promoter (5' site), whereas the region around the AP-1 sites remained unchanged¹⁶ (Fig. 2f). In agreement with previous observations, we detected substantial histone H3 acetylation in the 5' area of the promoter by ChIP, but no significant H3 acetylation around the AP-1 sites where c-Jun is localized (Fig. 2g, h). As expected, JNK inhibition induced a marked decrease in histone H3 acetylation,

but the regulation of H3 acetylation at the *c-jun* promoter was disrupted by Mbd3 depletion. H3 acetylation was slightly reduced in untreated cells, but in response to JNK inhibition no decrease in H3 acetylation was observed (Fig. 2h). Thus, Mbd3 contributes to the epigenetic regulation of the *c-jun* promoter by JNK signalling.

c-Jun has an important function in regulating intestinal epithelial homeostasis and intestinal cancer^{3,4}. Quantitative real-time polymerase chain reaction (qRT-PCR) analysis revealed an enrichment of *mbd3* mRNA in intestinal crypts (Supplementary Fig. 7a), and analysis of a knock-in mouse that expresses the reporter protein β -galactosidase under the control of the endogenous *mbd3* promoter (Supplementary Fig. 8a)¹⁷ confirmed high *mbd3* expression in the crypt (Supplementary Fig. 7b). To address a potential role of Mbd3 in controlling c-Jun function in the intestine, we generated a mouse line in which exon 1 of Mbd3 was flanked with two *loxP* sites (Supplementary Fig. 8a). Exon 1 of Mbd3 encodes the N-terminal portion of the MBD and after Cre-mediated recombination *mbd3a* and *mbd3b*, the two transcripts

generated from the *mbd3* locus, are not produced (Supplementary Fig. 8b, c). We crossed the Mbd3 floxed mouse line with a Villin-Cre line¹⁸ to specifically inactivate *mbd3* in the gut (*mbd3*^{ΔG/ΔG}). Deletion of the floxed *mbd3* allele was efficient (Supplementary Fig. 8d) and mRNA was isolated from crypts to investigate AP-1-dependent gene expression. Transcription of *c-jun* was increased in intestinal crypts lacking Mbd3, as was the transcription of several c-Jun target genes including *cd44* and *cyclind1* (also known as *Ccnd1*) (Fig. 3a). In addition, the expression of the stem cell marker *lgr5* was also increased in *mbd3*^{ΔG/ΔG} intestine (Fig. 3a) whereas *bmi1* remained unchanged. *lgr5* marks crypt base columnar (CBC) cells, which function as intestinal stem cells¹⁹. The increase in *lgr5* mRNA was not due to an increase of CBC cells as their number was normal in the absence of Mbd3 (Fig. 3b). c-Jun is highly expressed in CBC cells^{3,20}, and *lgr5* expression is reduced in crypts lacking c-Jun³. We thus analysed whether *lgr5* might be directly regulated by c-Jun/Mbd3. *In silico* analysis revealed the presence of a consensus AP-1 site in intron 1 of the *lgr5* gene (Fig. 3c) and ChIP analysis

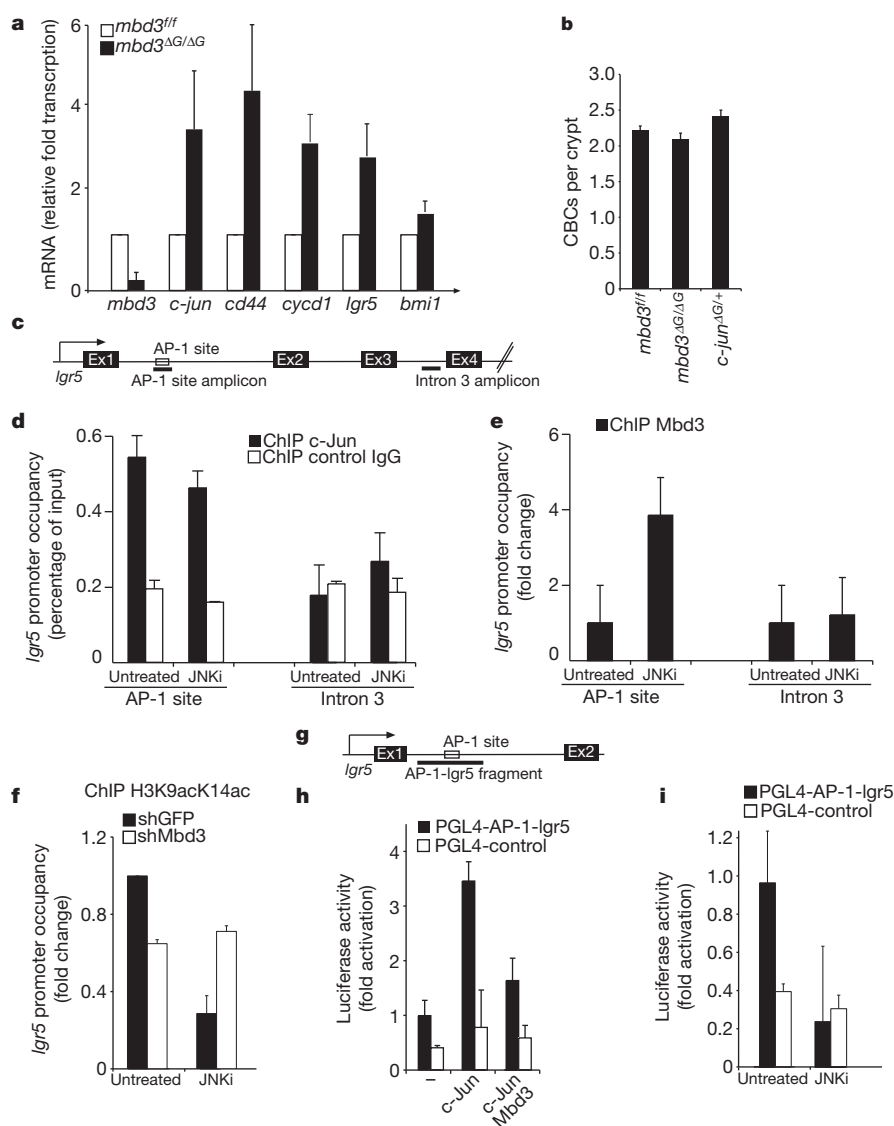


Figure 3 | *lgr5* is regulated by c-Jun-Mbd3/NuRD. **a**, qRT-PCR analysis of indicated genes on mRNA isolated from intestines of *mbd3*^{fl/fl} and *mbd3*^{ΔG/ΔG} mice ($n = 4$). Data were normalized to β -actin. **b**, Quantification of crypt base columnar cells (CBC) in *mbd3*^{fl/fl} ($n = 4$), *mbd3*^{ΔG/ΔG} ($n = 3$) and *mbd3*^{ΔG/ΔG} *c-jun*^{ΔG/+} ($n = 3$) mice. **c**, Schematic overview of the *lgr5* locus. Underlined regions show the amplification region of each primer pair used. **d**, ChIP for the *lgr5* AP-1 and intron 3 sites using c-Jun-specific or control antibody with or without JNKi. **e**, Mbd3 ChIP for the AP-1 and intron 3 sites in the *lgr5* locus.

f, Mbd3 ChIP for the *lgr5* AP-1 site with or without JNKi. **g**, Schematic overview of the fragment used for AP-1-lgr5 luciferase construct. **h**, **i**, Activity of an AP-1-lgr5 luciferase construct (PGL4-AP-1-lgr5) or empty vector (PGL4-control_Luc) relative to a tk-renilla luciferase transfection control in cells expressing c-Jun and Mbd3 (**h**) or with or without JNKi treatment (**i**). In all experiments (**c**–**f**, **h** and **i**) error bars show s.e.m. HCT116 cells were used and $n = 3$. JNKi treatment was for 30 min in **d**–**f** and for 5 h in **i**.

showed binding of c-Jun to this putative AP-1 site, but not to a region in intron 3 of the *lgr5* gene, used as a negative control (Fig. 3d). Mbd3 was also recruited to the same site in intron 1 of the *lgr5* gene, and JNKi increased Mbd3 recruitment (Fig. 3e). In addition, JNKi augmented CHD3 binding to the *lgr5* and *c-jun* AP-1 sites to a similar extent

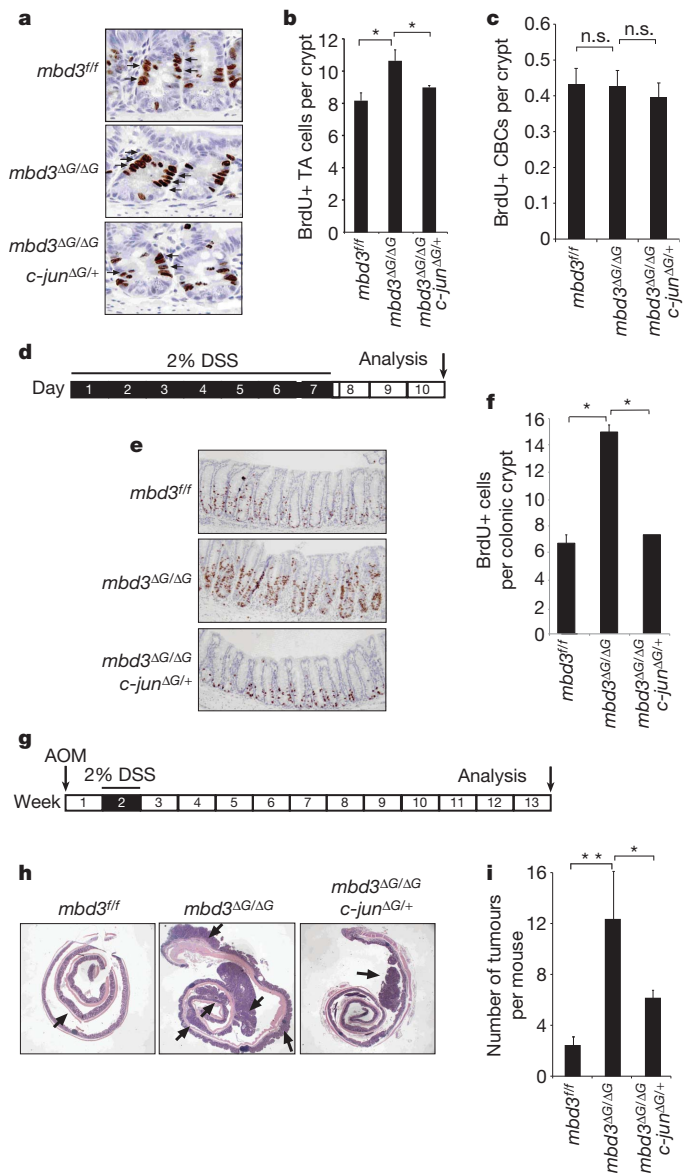


Figure 4 | Mbd3 antagonises c-Jun/AP-1 function in vivo.

a, Immunohistochemistry for BrdU+ cells on representative crypts of the indicated genotypes. Black arrows point to BrdU+ proliferative progenitors. **b**, Quantification of BrdU-positive (BrdU+) transit amplifying (TA) cells in *mbd3^{fl/fl}* ($n = 6$), *mbd3^{ΔG/ΔG}* ($n = 5$) and *mbd3^{ΔG/ΔG}; c-jun^{ΔG/+}* ($n = 4$) in intestinal crypts (error bars show s.e.m.; * $P = 0.001$; Student's t -test). **c**, Quantification of BrdU+ CBCs in *mbd3^{fl/fl}* ($n = 4$), *mbd3^{ΔG/ΔG}* ($n = 3$) and *mbd3^{ΔG/ΔG}; c-jun^{ΔG/+}* ($n = 3$) intestinal crypts. n.s., not significant. **d**, Schematic representation of the experimental design for DSS-induced intestinal regeneration. **e**, Immunohistochemistry for BrdU+ cells on representative colonic crypts of mice with the indicated genotypes. **f**, Quantification of BrdU+ cells in *mbd3^{fl/fl}* ($n = 10$), *mbd3^{ΔG/ΔG}* ($n = 6$) and *mbd3^{ΔG/ΔG}; c-jun^{ΔG/+}* ($n = 2$) mice colonic crypts; * $P = 0.00012$. **g**, Schematic representation of the experimental design for colitis-induced tumorigenesis. **h**, Haematoxylin and eosin staining on representative colons of mice with the indicated genotypes. **i**, Quantification of number of AOM/DSS-induced tumours in *mbd3^{fl/fl}* ($n = 5$), *mbd3^{ΔG/ΔG}* ($n = 4$) and *mbd3^{ΔG/ΔG}; c-jun^{ΔG/+}* ($n = 2$) mice (** $P = 0.01$, * $P = 0.016$). In all experiments error bars show s.e.m. and Student's t -test was used for statistical analysis.

(Supplementary Fig. 6e, f). JNK inhibition reduced acetylation of the *lgr5* AP-1 site but the Mbd3 depletion impaired JNKi-induced deacetylation (Fig. 3f). When inserted into a luciferase reporter construct, the *lgr5* fragment containing the AP-1 site (pGL4-AP-1-*lgr5*) (Fig. 3g) was activated by c-Jun overexpression whereas coexpression with Mbd3 resulted in repression (Fig. 3h). JNKi reduced pGL4-AP-1-*lgr5* reporter activity to background levels (Fig. 3i). These data suggest that Mbd3/c-Jun regulates *lgr5* transcription.

The only described biological function of Mbd3 is in controlling proliferation and differentiation of embryonic stem cells^{21,22}, so we investigated whether intestinal cell lineage specification was impaired in the gut of *mbd3^{ΔG/ΔG}* mice. Immunohistochemical stainings for Alcian blue/periodic acid Schiff (AB/PAS) (which detects goblet cells), chromogranin (for enteroendocrine cells) or lysozyme (for Paneth cells) failed to detect a defect in cell fate specification in the intestine of *mbd3^{ΔG/ΔG}* mice (Supplementary Fig. 9a–c). Interestingly, *mbd3^{ΔG/ΔG}* mice showed a significant increase in bromodeoxyuridine (BrdU)-positive transient amplifying cells in the crypts compared to controls (Fig. 4a, b), but we did not detect any increase in CBC cell proliferation (Fig. 4c). To test whether increased c-Jun activity was responsible for the augmented intestinal progenitor proliferation in *mbd3^{ΔG/ΔG}* mice, one *c-jun* allele was deleted in a *mbd3^{ΔG/ΔG}* background, generating *mbd3^{ΔG/ΔG}; c-jun^{ΔG/+}* mice. The hyperproliferation phenotype in the absence of Mbd3 was rescued in *mbd3^{ΔG/ΔG}; c-jun^{ΔG/+}* mice (Fig. 4a, b). We next investigated the impact of *mbd3* inactivation under pathological conditions. Intestinal colitis and regeneration was induced by dextran sodium sulphate (DSS) administration (Fig. 4d). *mbd3^{ΔG/ΔG}* mice showed a large increase in proliferating cells in the colon after DSS treatment compared to control animals. The hyperproliferation phenotype was completely reversed by deletion of a single *c-jun* allele (Fig. 4e, f). Moreover, to investigate a potential role of Mbd3 in intestinal tumorigenesis, we applied the azoxymethane (AOM)/DSS protocol to trigger colitis-induced tumorigenesis (Fig. 4g). Mbd3 deletion resulted in a striking increase in tumour number (Fig. 4h, i) as well as an increase in tumour size (Supplementary Fig. 9d) that was significantly rescued by concomitant deletion of one *c-jun* allele (Fig. 4h, i and Supplementary Fig. 9d). Thus, in the intestine Mbd3 controls cellular proliferation under physiological and pathological conditions by antagonising c-Jun activity.

The direct recruitment of the NuRD complex by Mbd3 interaction with unphosphorylated c-Jun suggests that in the absence of JNK pathway activation, c-Jun actively represses AP-1 target gene expression, providing a molecular explanation for previous observations that c-Jun phosphorylation mediates dissociation of an inhibitory complex associated with histone deacetylation^{23,24}. The phosphorylation-dependent interaction of Mbd3/c-Jun is in agreement with a three-step mechanism of c-Jun activation by JNK. Upon JNK activation, NuRD dissociates from c-Jun, resulting in derepression of target gene transcription. Subsequently, a coactivator protein would bind to the c-Jun TAD and further augment gene expression. It is worth noting that to explain stimulation of c-Jun activity by N-terminal phosphorylation, coactivator binding may, but is not required to be phosphorylation-dependent^{4,25}. Lastly, upon termination of the JNK signal, Mbd3/NuRD would re-bind to unphosphorylated c-Jun resulting in cessation of target gene expression. Thus, the TAD of c-Jun functions as a repressor domain that is negatively regulated by phosphorylation.

METHODS SUMMARY

Cell culture. HCT116 and HEK293T cells were grown in DMEM and 10% FBS. **IP, pull-down and western blot analysis.** Cells were lysed in the appropriate buffer for immunoprecipitation, GST pull-down or peptide pull-down assays. Immunoblots were carried out as previously described⁴.

Reporter gene assay. Firefly and *Renilla* luciferase reporter assays were performed and measured using the Dual-Luciferase Reporter Assay System (Promega).

Chromatin immunoprecipitation (ChIP). ChIP analysis was performed as described previously²⁶. qRT-PCR was accomplished with SYBR Green incorporation.

Mouse lines. Mbd3^{AG/AG} mice resulted from crossing Mbd3 floxed line²¹ with villin cre line¹⁸.

AOM/DSS model of colon carcinoma and DSS model of colon inflammation. The AOM/DSS model²⁷ and the DSS protocol used²⁸ have been previously described. For tissue collection, mice were sacrificed and the intestines were frozen in liquid nitrogen. For qRT-PCR analysis, total mRNA was isolated from dissected intestines, or from villus and crypt fractions obtained as described before¹⁹.

Immunohistochemical staining. For haematoxylin & eosin staining and immunohistochemistry, the intestines were fixed overnight in 10% neutral buffered formalin, transferred into 70% ethanol, rolled, processed and embedded into paraffin. Sections were cut at 4 µm.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 September; accepted 25 October 2010.

Published online 2 January 2011.

- Davis, R. J. Signal transduction by the JNK group of MAP kinases. *Cell* **103**, 239–252 (2000).
- Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nature Rev. Cancer* **3**, 859–868 (2003).
- Sancho, R. *et al.* JNK signalling modulates intestinal homeostasis and tumorigenesis in mice. *EMBO J.* **28**, 1843–1854 (2009).
- Nateri, A. S., Spencer-Dene, B. & Behrens, A. Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development. *Nature* **437**, 281–285 (2005).
- Pulverer, B. J., Kyriakis, J. M., Avruch, J., Nikolakaki, E. & Woodgett, J. R. Phosphorylation of c-jun mediated by MAP kinases. *Nature* **353**, 670–674 (1991).
- Behrens, A., Sibilio, M. & Wagner, E. F. Amino-terminal phosphorylation of c-Jun regulates stress-induced apoptosis and cellular proliferation. *Nature Genet.* **21**, 326–329 (1999).
- Nateri, A. S., Riera-Sans, L., Da Costa, C. & Behrens, A. The ubiquitin ligase SCFFbw7 antagonizes apoptotic JNK signaling. *Science* **303**, 1374–1378 (2004).
- Le Guezennec, X. *et al.* MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* **26**, 843–851 (2006).
- Denslow, S. A. & Wade, P. A. The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26**, 5433–5438 (2007).
- Ahringer, J. NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet.* **16**, 351–356 (2000).
- Bowen, N. J., Fujita, N., Kajita, M. & Wade, P. A. Mi-2/NuRD: multiple complexes for many purposes. *Biochim. Biophys. Acta* **1677**, 52–57 (2004).
- Saito, M. & Ishikawa, F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.* **277**, 35434–35439 (2002).
- Hendrich, B. & Tweedie, S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* **19**, 269–277 (2003).
- D'Orazio, D. *et al.* Cooperation of two PEA3/AP1 sites in uPA gene induction by TPA and FGF-2. *Gene* **201**, 179–187 (1997).
- Angel, P., Hattori, K., Smeal, T. & Karin, M. The *jun* proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell* **55**, 875–885 (1988).
- Clayton, A. L., Rose, S., Barratt, M. J. & Mahadevan, L. C. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *EMBO J.* **19**, 3714–3726 (2000).
- Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* **15**, 710–723 (2001).
- El Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
- van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
- Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature Cell Biol.* **8**, 285–292 (2006).
- Kaji, K., Nichols, J. & Hendrich, B. Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* **134**, 1123–1132 (2007).
- Baichwal, V. R., Park, A. & Tjian, R. The cell-type-specific activator region of c-Jun juxtaposes constitutive and negatively regulated domains. *Genes Dev.* **6**, 1493–1502 (1992).
- Weiss, C. *et al.* JNK phosphorylation relieves HDAC3-dependent suppression of the transcriptional activity of c-Jun. *EMBO J.* **22**, 3686–3695 (2003).
- Arias, J. *et al.* Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature* **370**, 226–229 (1994).
- Nelson, J. D., Denisenko, O. & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature Protocols* **1**, 179–185 (2006).
- Neufert, C., Becker, C. & Neurath, M. F. An inducible mouse model of colon carcinogenesis for the analysis of sporadic and inflammation-driven tumor progression. *Nature Protocols* **2**, 1998–2004 (2007).
- Floer, M. *et al.* Enoxaparin improves the course of dextran sodium sulfate-induced colitis in syndecan-1-deficient mice. *Am. J. Pathol.* **176**, 146–157 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the LRI Animal Unit, Equipment Park, FACS, Peptide synthesis and the Experimental Histopathology unit for technical help and O. S. Gabrielsen for reagents. We thank C. Hill and H. Van Dam for critical reading of the manuscript. C.A. and R.S. were funded by Marie Curie Intra-European Fellowships (PIEF-GA-2008-220566 and MEIF-CT-2006-041119). The London Research Institute is funded by Cancer Research UK.

Author Contributions C.A. designed and performed most of the experiments, analysed data, and co-wrote the paper. K.N. identified Mbd3 as binding specifically to unphosphorylated c-Jun using the yeast three-hybrid screen, generated reagents and provided Fig. 1a, c and Supplementary Fig. 3. R.S. helped with all *in vivo* experiments and provided Fig. 3a and Supplementary Fig. 7a. A.C. provided Fig. 1f–i. B.H. generated the *mbd3* floxed mouse line. A.B. supervised all aspects of this work and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to axel.behrens@cancer.org.uk

METHODS

Cell culture and transfection. HCT116 and HEK293T cells were cultured in Dulbecco's modified Eagle medium (DMEM) and 10% FBS. Cells were plated at subconfluence and transfected with lipofectamine reagent (Invitrogen). HCT116 cells were treated when indicated with 50 μ M JNKi (SP600125, Calbiochem) or 25 ng ml⁻¹ anisomycin (Sigma).

Immunoprecipitation. Cells were lysed for 30 min at 4 °C in 500 μ l of immunoprecipitation buffer containing 0.5% Triton X-100, 1 mM EDTA, 100 μ M Na-orthovanadate, 0.25 mM PMSF (phenylmethylsulphonyl fluoride), and protease inhibitor mixture (Sigma) in PBS. After centrifugation, supernatants were incubated for 3 h at 4 °C with 1 μ g of the anti-Flag (Sigma) antibody coupled to protein A-Sepharose beads.

GST pull-down. GST fusion proteins were purified from the *Escherichia coli* BL-21 strain in a buffer containing 20 mM Tris-HCl, pH 7.4, 1 M NaCl, 0.2 mM EDTA, 1 mM dithiothreitol (DTT), 1 mg ml⁻¹ lysozyme, 1 mM PMSF, and protease inhibitor complex (Sigma) and bound to glutathione-Sepharose (Amersham Biosciences). GST fusion proteins were purified and incubated with 400 μ g of cell lysates for 2 h at 4 °C in RIPA buffer and extensively washed. Pulled down proteins were analysed by western blot.

Peptide pull-down. Biotinylated peptide (1 mg) was incubated with Dynabeads M280 streptavidin in buffer A (5 mM Tris-Cl pH 7.5, 0.5 mM EDTA, 1 M NaCl) at 4 °C overnight. Peptide bound beads were washed and resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.2% NP 40, 1 mM DTT). 293T cells overexpressing c-Jun were lysed in lysis buffer and then incubated with different Dynabead-bound peptides for 2 h at 4 °C. The beads were washed three times with lysis buffer containing 400 mM NaCl. Pulled-down proteins were analysed by western blot.

Western blot analysis. HCT116 cells lysates were homogenized in RIPA lysis buffer supplemented with protease inhibitor (Sigma). Immunoblots were carried out as described previously⁴. The gels were transferred to nitrocellulose membranes, and the membranes were immunoblotted with various antibodies as indicated: anti-c-Jun (H79, Santa Cruz sc1694 and BD Transduction Laboratories 610326), anti-Myc (9E10 clone), anti-Flag-HRP (horseradish peroxidase-conjugated; Sigma), anti-c-Jun-ser63 and anti- β -Actin (Sigma), anti-Mbd3 (c-18, Santa Cruz sc-9402), anti-JNK1/2 (NEB cell signalling 9252), anti-CHD3 (BD Transduction Laboratories 611847) and anti-MTA3 (Santa Cruz sc-48799). HCT116 cells were treated where indicated with ultraviolet at 40 J cm⁻².

Reporter gene assay. HCT116 cells were transfected with the indicated plasmids with lipofectamine reagent (Invitrogen). Transient transfections of the indicated plasmids and controls of firefly and *Renilla* luciferase reporters were performed and measured using the Dual-Luciferase Reporter Assay System (Promega), 36 h post-transfection. Data are expressed as fold induction after being normalized using tk-*renilla* luciferase (mean \pm s.e.m.; $n = 3$).

Chromatin immunoprecipitation (ChIP). ChIP analysis was performed as described previously²⁶. Cells were treated as indicated before collection, with 50 μ M JNKi (SP600125, Calbiochem), 25 ng ml⁻¹ anisomycin (Sigma) or transfected with the indicated plasmids with lipofectamine reagent (Invitrogen). Immunoprecipitations were carried out with the indicated antibodies, anti-Mbd3 (c-18) (Santa Cruz sc9402), anti-HDAC1 (H51) (Santa Cruz sc7872), anti-c-Jun (H79) (Santa Cruz sc1694) or anti-acetyl K9, K14 histone 3 (Upstate 06-599) and collected by agarose beads (Santa-Cruz). The oligonucleotide sequences used to amplify the DNA fragments by qRT-PCR are AP1 c-Jun F, 5'-GCGACGCGAGCCAATG-3'; AP1 c-Jun R, 5'-AGCCGAGCTCAACAC TTATCT-3'; 5' c-Jun F, 5'-TTTACTCTCTCTCCCCAGCTA-3'; 5' c-Jun R, 5'-CCTCCAGCCTCTGAAACATC-3'; Igr5-AP-1 F, 5'-TGGAATCA GTGCGGTTC-3'; Igr5-AP-1 R, 5'-CAGGGAAATGGGCAACAAG-3'; Igr5-intron3 F, 5'-TCTGCCTCAGGCTTACATGGA-3'; Igr5-intron3 R, 5'-CA CAAGAATTCTGCAGCACATTT-3'; F-CD44 AP-1, 5'-AGTGCAGTGGCAGC ATCTTG-3'; R-CD44 AP-1, 5'-CGGGAGAAATCGCTTGAACAC-3'.

Quantitative real-time PCR was performed with SYBR Green incorporation (Platinum Quantitative PCR SuperMix-UDG w/ROX, Invitrogen) using an ABI7900HT (Applied Bioscience) and the data were analysed using the SDS 2.3 software.

RNA extraction and qRT-PCR analysis. Total mRNA was isolated using RNeasy Mini-kit according to the manufacturer's instructions (Qiagen). cDNA was synthesized using Invitrogen Superscript reagents according to the manufacturer's instructions. The oligonucleotide sequences used to amplify the DNA fragments by qRT-PCR are F-h-c-Jun, 5'-TCGACATGGAGTCCCAGGA-3'; R-h-c-Jun,

5'-GGCGATTCTCTCCAGCTTCC-3'; F-h-actin, 5'-GGATGCAGAAGGAG ATCACTG-3'; R-h-actin, 5'-CGATCCACACGGAGTACTTG-3'; F-hCD44, 5'-CTCCTGGCACTGGCTCTGA-3'; R-hCD44, 5'-CTGCCCCACACCTTCT CCTACTATT-3'; F-mMbd3, 5'-CACCGCGTGCCTGTAC-3'; R-mMbd3, 5'-TGGTACCGGTTGCTTGA-3'; F-mc-Jun, 5'-TGAAAGCTGTGTCCTC TGTC-3'; R-mc-Jun, 5'-ATCACAGCACATGCCACTTC-3'; F-m β -actin, 5'-ATGCTCCCCGGGTGTAT-3'; R-m β -actin, 5'-CATAGGAGTCTTCTG ACCCAATTC-3'; F-mCD44, 5'-CTCCTGGCACTGGCTCTGA-3'; R-mCD44, 5'-CTGCCCCACACCTTCTCTACTATT-3'; F-mCCDN1, 5'-GTGCGTGCA GAAGGAGATTGT-3'; R-mCCDN1, 5'-CTCACAGACCTCCAGCATCCA-3'; F-mLgr5, 5'-CGGAGGAAGCGCTACAGAAT-3'; R-mLgr5, 5'-CTGGGT GGCACGTAGCTGAT-3'; F-mBmiI, 5'-GGGCTTTTCAAAAATGAGATGAA-3'; R-mBmiI, 5'-GAGCCATTGGCAGCATCAG-3'.

Quantitative real-time PCR was accomplished with SYBR Green incorporation (Platinum Quantitative PCR SuperMix-UDG w/ROX, Invitrogen) using an ABI7900HT (Applied Bioscience), and the data were analysed using the SDS 2.3 software. Results were normalized to those obtained with β -actin.

RNAi constructs. siRNA sequences directed towards the coding region of the indicated genes were designed using the Dharmacon siDESIGN centre. Annealed hairpin oligonucleotides were cloned into pSuper (OligoEngine).

shGFP, sense strand 5'-GAACGGCATCAAGGTGAAC-3'; shJNK1/2, sense strand 5'-AAAGAATGTCCTACCTTCTT-3'; shMbd3-a, sense strand 5'-AGA CGGCGTCCATCTTCAA-3'; shMbd3-b, sense strand 5'-TATGGTCAAC ACCACGCA-3'; shMbd3-c, sense strand 5'-GGACATCAGGAAGCAGAA-3'.

Mouse lines. Embryonic stem cells in which one *mbd3* allele is deleted and the other contains *loxP* sites flanking exon 1 have been described²¹. Upon transfection with Cre recombinase, the resulting *mbd3* *Ex1*^{Δ/-} embryonic stem cells lack Mbd3a and Mbd3b. *mbd3*^{Fllox/-} embryonic stem cells were used for blastocyst injection to generate a mouse line homozygous for the *mbd3*^{Ex1-Flox} allele. These mice were viable and fertile. The Villin-Cre line has been described previously¹⁸. For tissue collection, mice were killed by cervical dislocation and the intestines were removed and flushed extensively with cold PBS and frozen in liquid nitrogen. For quantitative real-time PCR (qRT-PCR) analysis, total mRNA was isolated from dissected intestines, or from villus and crypt fractions obtained as described before¹⁹. Results were normalized to those obtained with β -actin and results are presented as fold induction over control mice.

AOM/DSS model of colon carcinoma. The AOM/DSS model used has been described previously²⁷. Mice were injected intraperitoneally (i.p.) with 10 mg kg⁻¹ body weight of AOM (Sigma) dissolved in physiological saline. Seven days later, 2% DSS was given in the drinking water over 7 days, followed by normal water until the end of the experiment. Body weight was measured every week and the animals were killed 13 weeks after AOM injection for histological analysis. The number of tumours, incidence of tumours and tumour area were determined. For biochemical characterization dissected tumours were snap-frozen in liquid nitrogen and used for qRT-PCR analysis.

DSS model of colon inflammation. The DSS model used has been described previously²⁸. Mice were administered orally with 2% DSS in drinking water starting on day 1 and continued until day 6. On day 7, the water was changed to normal drinking water. Body weight was measured every day, and the animals were killed at day 10 for histological analysis.

Immunohistochemical staining. Mice were injected i.p. with 100 mg kg⁻¹ BrdU (Sigma) 1.5 h before killing. Mice were euthanized by cervical dislocation, the small intestines dissected out into ice-cold PBS and the faecal contents flushed out with ice-cold PBS. The intestines were cut longitudinally into pieces of similar size, opened out and fixed overnight in 10% neutral buffered formalin, briefly washed with PBS and transferred into 70% ethanol, rolled, processed and embedded into paraffin. Sections were cut at 4 μ m for haematoxylin and eosin staining. For immunohistochemistry, antibodies against BrdU (BD Biosciences) were used. To quantify the BrdU-positive cells per crypt, 100 full crypts were scored. For AB/PAS staining, immunohistochemistry and immunofluorescence antibodies against chromogranin-A (Abcam) and β -galactosidase (Acris Antibodies) were used. Anti-lysozyme (DAKO) was used for immunofluorescence as described before⁴. To quantify the AB/PAS⁺ or chromogranin⁺ cells per villus, 100 full villi were scored.

Statistics. Statistical evaluation was performed using the Student's unpaired *t*-test. Data are presented as mean \pm s.e.m. and $P \leq 0.05$ was considered statistically significant.

Structure and function of an irreversible agonist- β_2 adrenoceptor complex

Daniel M. Rosenbaum^{1†}, Cheng Zhang¹, Joseph A. Lyons^{2,3}, Ralph Holl⁴, David Aragao³, Daniel H. Arlow⁵, Søren G. F. Rasmussen¹, Hee-Jung Choi^{1,6}, Brian T. DeVree⁷, Roger K. Sunahara⁷, Pil Seok Chae⁸, Samuel H. Gellman⁸, Ron O. Dror⁵, David E. Shaw⁵, William I. Weis^{1,6}, Martin Caffrey³, Peter Gmeiner⁴ & Brian K. Kobilka¹

G-protein-coupled receptors (GPCRs) are eukaryotic integral membrane proteins that modulate biological function by initiating cellular signalling in response to chemically diverse agonists. Despite recent progress in the structural biology of GPCRs¹, the molecular basis for agonist binding and allosteric modulation of these proteins is poorly understood. Structural knowledge of agonist-bound states is essential for deciphering the mechanism of receptor activation, and for structure-guided design and optimization of ligands. However, the crystallization of agonist-bound GPCRs has been hampered by modest affinities and rapid off-rates of available agonists. Using the inactive structure of the human β_2 adrenergic receptor (β_2 AR) as a guide, we designed a β_2 AR agonist that can be covalently tethered to a specific site on the receptor through a disulphide bond. The covalent β_2 AR-agonist complex forms efficiently, and is capable of activating a heterotrimeric G protein. We crystallized a covalent agonist-bound β_2 AR-T4L fusion protein in lipid bilayers through the use of the lipidic mesophase method², and determined its structure at 3.5 Å resolution. A comparison to the inactive structure and an antibody-stabilized active structure (companion paper³) shows how binding events at both the extracellular and intracellular surfaces are required to stabilize an active conformation of the receptor. The structures are in agreement with long-timescale (up to 30 μ s) molecular dynamics simulations showing that an agonist-bound active conformation spontaneously relaxes to an inactive-like conformation in the absence of a G protein or stabilizing antibody.

The relationship between agonist binding to GPCRs and the conformational changes that facilitate G protein binding and activation remains largely unknown. Characterization of GPCR activation at a molecular level has been driven by a combination of X-ray crystal structure analysis and spectroscopic approaches. Rhodopsin has served as a prototype GPCR for biophysical studies, given its ready availability and superior stability. Crystal structures have been obtained for the inactive dark state with inverse agonist 11-*cis*-retinal covalently bound^{4,5}, as well as the active state mimetic low-pH opsin lacking the retinal chromophore^{6,7}. Opsin differs from rhodopsin by outward rigid-body movements of transmembrane helices (TMs) 5 and 6 at the cytoplasmic G protein binding surface. For other GPCRs, which respond to diffusible ligands, structural information has proven more difficult to obtain. Fluorescence spectroscopy studies show that activation of the β_2 AR by diffusible ligands can follow multiple pathways, with a complex energy landscape of receptor conformations⁸. This multitude of accessible conformations has probably contributed to the difficulty in obtaining crystal structures of non-rhodopsin GPCRs.

Inactive state crystal structures of the β_2 AR, β_1 AR, and A_{2A} adenosine receptor have been solved over the past several years with the aid of protein engineering techniques¹, but agonist-bound structures for these proteins have yet to be reported. Like many GPCRs, the β_2 AR has two agonist affinity states: a low affinity state in the absence of cognate G protein, and a high affinity state in the presence of G protein. This observation indicates that agonists can bind to two distinct receptor conformations. A more complete understanding of the processes of agonist binding and activation requires structures of both high and low affinity states. The high affinity state is challenging because a receptor G protein complex is unstable in detergent solutions required for purification of both GPCRs and heterotrimeric G proteins. In a companion paper³, we describe the use of a conformationally selective camelid antibody (nanobody, Nb80) with G-protein-like properties to obtain a structure of a high-affinity agonist-bound conformation. Obtaining a structure of the low affinity state is also challenging because of the relatively rapid association and dissociation rates of commercial β_2 AR agonists. Inspired by the covalent retinal-rhodopsin system, we proposed that the ability to crystallize an agonist-bound GPCR would be enhanced by chemically crosslinking ligand and receptor, preferably in a manner that would not inhibit conformational freedom and the capacity to activate a G protein.

Our design strategy for a covalent β_2 AR agonist was to combine a β -adrenergic agonist core (procaterol) and a reactive chemical group that could be targeted to a specific residue on the receptor. Using the structure of the carazolol-bound β_2 AR as a template (Fig. 1a), a flexible linker was added to bridge these two components such that the covalent attachment would not inhibit binding of the agonist core or conformational flexibility of the transmembrane helices. Biochemical precedent for this strategy came from the covalent labelling reagent BABC (Fig. 1b), in which an electrophilic group appended to the carazolol ligand core was determined to react with His 93^{2,64} at the extracellular end of TM2 (ref. 9; Ballesteros-Weinstein numbering¹⁰ used in superscript). For crosslinking, we chose the reaction between a free cysteine on the receptor (introduced at position 93) and a ligand disulphide moiety, based on the mild and proximity-dependent 'tethering' approach that has proven broadly applicable to different protein targets including GPCRs¹¹. The designed covalent β agonist FAUC50 (Fig. 1b) was synthesized in enantiomerically pure form^{12,13}, along with the noncovalent analogue FAUC72 (Fig. 1b, Supplementary Fig. 1 and Supplementary Information) for use in control experiments.

Incubation of compound FAUC50 with mutant H93C receptor led to efficient and irreversible blocking of radioligand binding (Supplementary Fig. 2, Supplementary Information). We sought to determine whether the tethered FAUC50- β_2 AR^{H93C} complex is capable of activating a G protein. Wild-type and mutant receptor were

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 279 Campus Drive, Stanford, California 94305, USA. ²Department of Chemical and Environmental Sciences, University of Limerick, Limerick, Ireland. ³Membrane Structural and Functional Biology Group, School of Biochemistry and Immunology, Trinity College, Dublin 2, Ireland. ⁴Department of Chemistry and Pharmacy, Friedrich Alexander University, Schuhstrasse 19, 91052 Erlangen, Germany. ⁵D. E. Shaw Research, New York, New York 10036, USA. ⁶Department of Structural Biology, Stanford University School of Medicine, 299 Campus Drive, Stanford, California 94305, USA. ⁷Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ⁸Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA. [†]Present address: Department of Biochemistry, University of Texas Southwestern, 5223 Harry Hines Blvd, Dallas, Texas 75390-8816, USA.

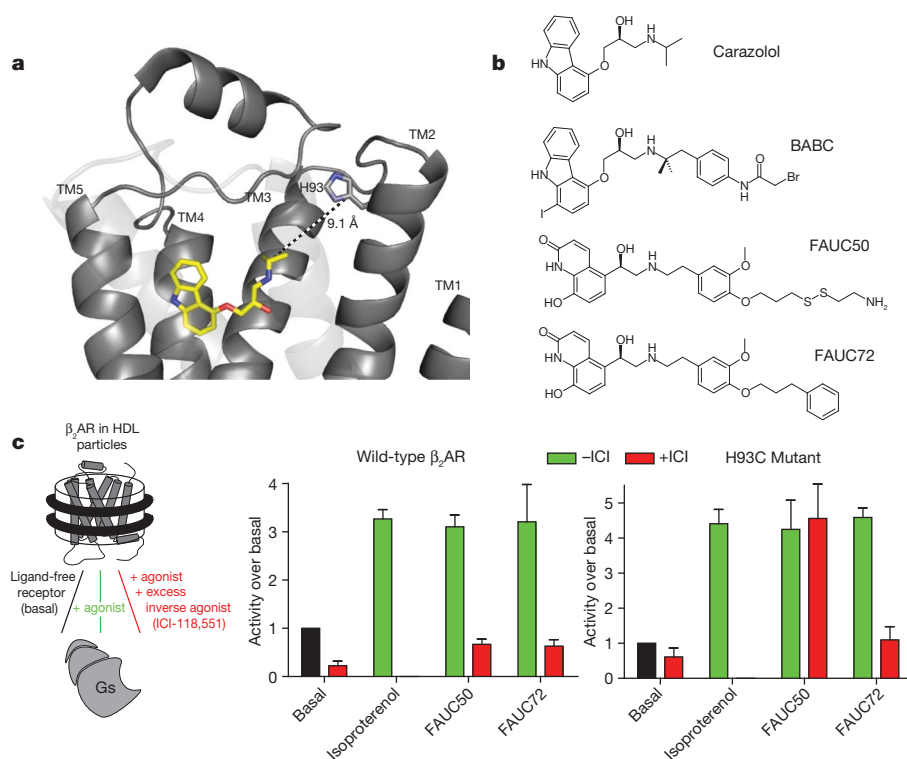


Figure 1 | Design and function of a covalent agonist. **a**, Structure of the carazolol-bound β_2 AR, receptor in grey cartoon and ligand in yellow sticks, showing distance between isopropyl group and His 93^{2,64} imidazole. **b**, Structures of carazolol and the related BABC ligand, covalent ligand

reconstituted into high density lipoprotein (HDL) particles¹⁴, and then incubated with ligand alone or ligand followed by an excess of the high-affinity inverse agonist ICI-118,551. Heterotrimeric Gs protein was then added to the particles, and activation was observed by measuring GTP γ S binding to the Gs α subunit. Figure 1c shows that the inverse agonist treatment prevented agonist-induced G protein activation by the wild-type receptor. However, excess ICI-118,551 is unable to reverse FAUC50-induced coupling in the case of β_2 AR^{H93C}. The noncovalent analogue FAUC72 is displaced by the inverse agonist for both wild-type and H93C mutant receptors. From these experiments we conclude that FAUC50 not only reacts efficiently with Cys 93 to form a covalent complex, but this complex is also capable of activating a G protein.

We were motivated to develop a covalent agonist by our inability to produce diffraction-quality crystals of the previously described β_2 AR-T4 lysozyme chimaera (β_2 AR-T4L) in complex with available non-covalent agonists. In contrast, the purified FAUC50- β_2 AR^{H93C}T4L complex readily yielded diffraction-quality crystals in lipidic mesophases¹⁵. We overexpressed the mutant receptor in Sf9 insect cells, purified the protein to homogeneity using immunoaffinity and ligand-affinity chromatography, and introduced the FAUC50 ligand during a subsequent chromatography step. Using a cholesterol-doped monoolein cubic phase and robotic *in meso* technology¹⁶, we obtained $50 \times 15 \times 5 \mu\text{m}$ blade-shaped crystals (Supplementary Fig. 3) that diffracted to 3.5 Å. Combining diffraction data from 19 microcrystals (Supplementary Table 1), we solved the agonist-bound structure by molecular replacement using the coordinates of carazolol-bound β_2 AR-T4L¹⁷ as a search model. The packing of protomers in the orthorhombic crystals is distinct from previous crystals of β_2 AR-T4L fusion proteins. However, the orientation of T4L relative to the receptor closely resembles the original carazolol-bound structure (PDB ID 2RH1). An omit map at the ligand binding site reveals clear electron density for the agonist (Supplementary Fig. 4), and the refined structure shows prominent features expected from adrenergic receptor

FAUC50 and noncovalent analogue FAUC72. **c**, G protein activation assay demonstrating that covalently bound FAUC50 activates the β_2 AR. Error bars represent the standard deviation from three independent experiments.

pharmacology¹⁸: the amine and β -hydroxyl of FAUC50 contact Asp 113^{3,32} and Asn 312^{7,39} in a manner similar to inverse agonists¹⁷, whereas the aromatic moiety of FAUC50 that replaces the catechol ring forms hydrogen bonds with Ser 203^{5,42} and Ser 207^{5,46} (Fig. 2a, left).

The crystal structure of the covalent agonist-bound receptor in a lipidic mesophase does not show the conformational changes near the cytoplasmic surface associated with G protein binding, as observed in the nanobody-stabilized agonist complex (Fig. 2b). This was surprising given that the covalent agonist-bound receptor activates Gs (Fig. 1c) and the T4L fusion does not interfere with agonist-induced conformational changes detected with a fluorescent probe at the end of TM6 (ref. 17). This result implies that even an agonist with zero dissociation inefficiently stabilizes an active conformation of the β_2 AR in a lipid bilayer environment, consistent with a higher stability of the inactive carazolol-bound conformation. Nevertheless, allosteric communication between the ligand-binding pocket and the cytoplasmic G-protein-binding surface is a fundamental feature of GPCRs, exemplified by agonist-induced G protein stimulation and G-protein-induced high-affinity agonist binding¹⁴. For the β_2 AR, we can now compare an agonist-bound structure to that of a nanobody-stabilized active state mimetic (see companion paper³), to better understand conformational changes associated with activation. The superposition in Fig. 2b shows that agonist binding alone is insufficient to stabilize an active conformation at the cytoplasmic surface, with the requisite outward movement of TM5 and TM6. The largest differences proximal to the ligand-binding pocket in the nanobody-stabilized active conformation involve movement of Ile 121^{3,40} away from Pro 211^{5,50} and into space occupied by Phe 282^{6,44} in the inactive state (Fig. 2c, middle). The concomitant outward movement of Phe 282^{6,44} is accompanied by subtle backbone torsion changes distributed through TM6 that cause the cytoplasmic half of TM6 to be redirected outward, similar to its orientation in activated opsin^{6,7}. In the binding pocket, the most significant difference between the two agonist-bound structures and the carazolol-bound β_2 AR are in the hydrogen bonding contacts with

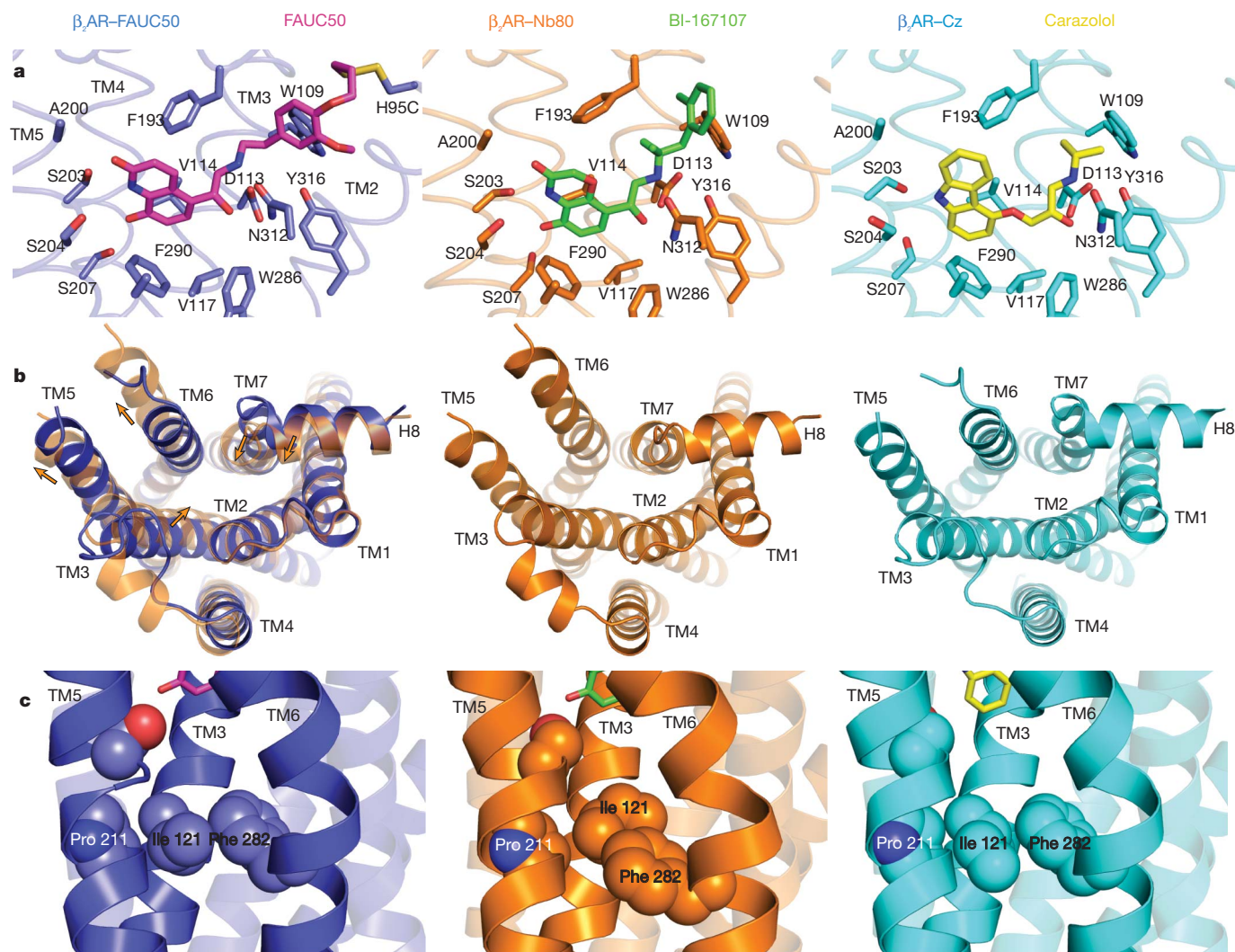


Figure 2 | Comparison of agonist and inverse agonist bound β_2 AR structures. Comparison of the covalent FAUC50-bound β_2 AR^{H93C}T4L (left panels, blue cartoon, ligand carbons in purple), BI-167107-bound β_2 AR-T4L/nanobody (β_2 AR-Nb80) complex (middle panels, orange cartoon, ligand carbons in green), and carazolol-bound β_2 AR-T4L (right panels, cyan cartoon, ligand carbons in yellow). **a**, Hormone binding site with interactions between ligands and receptors. TMs 6 and 7 and residues Phe 289, Asn 293 and Tyr 308

Ser 203^{5,42} and Ser 207^{5,46} on TM5 (Fig. 2a). However, only in the nanobody-stabilized structure are these differences coordinated with further changes towards the cytoplasmic surface of the molecule (Fig. 2b and c, middle).

To determine whether an agonist-bound receptor would sustain a stable active conformation in the absence of a cytoplasmic binding partner, we initiated an unbiased molecular dynamics (MD) simulation of the receptor from the nanobody complex structure, but with the nanobody removed (Fig. 3a). In the first several microseconds of simulated time, the intracellular ends of TM5 and TM6 showed high mobility (Supplementary Fig. 5), drifting by as much as 5 Å from their crystallographic positions. After approximately 11 μ s, the agonist-bound receptor spontaneously transitioned to a more rigid conformation resembling the inactive, carazolol-bound structure and the covalent agonist-bound structure (superposition in Fig. 3b), which remained stable for the remainder of the 30- μ s simulation (Fig. 3a, top, and Supplementary Fig. 6). In particular, TM5 and TM6 reverted to the positions they adopt in the inactive structure, as did a number of side chains, including Ile 121^{3,40} and Phe 282^{6,44} (Fig. 3a, bottom). Following the transition, the TM3–TM6 ionic lock was usually intact, as observed previously in simulations initiated from the inactive structure¹⁹.

are omitted for clarity. **b**, Comparison of the cytoplasmic faces showing differences in TMs 5 and 6. Superimposed β_2 AR-Nb80 complex is also shown in left panel as a transparent cartoon, and arrows indicate rigid body movements. **c**, Conformational switch region with residues Ile 121^{3,40}, Pro 211^{5,50} and Phe 282^{6,44} from TMs 3, 5 and 6 (other TMs transparent). Side chains are shown in van der Waals sphere representation.

Although this 30- μ s simulation is an order of magnitude longer than any previously published atomistic simulation of a membrane protein, the transition to an inactive-like conformation took place more quickly than the millisecond timescales observed experimentally for adrenergic receptor activation²⁰. An additional simulation with different protonation states for Asp 79^{2,50} and Asp 130^{3,49}—two conserved residues whose protonation states have been suggested to change upon receptor activation^{21,22}—showed similar behaviour (Supplementary Fig. 7).

In the dynamic conformational equilibrium of a GPCR that links diffusible agonist binding and G protein association, the energies of different states reflect both the ligand binding energies and the conformations of the receptor and its binding partners. This can be depicted in a hypothetical energy landscape of the β_2 AR as shown in Fig. 3c, where R, R', R'' and R* represent members of the ensemble of receptor conformations along an activation pathway. The constitutive activity displayed by the β_2 AR implies that the energy differences and barriers between inactive (R) and active (R*) conformations are low enough to allow a significant population of active state receptors even in the absence of agonist. Agonist binding decreases the energy difference and thus increases the population of receptors in an active conformation; however, the inactive state is still the most stable

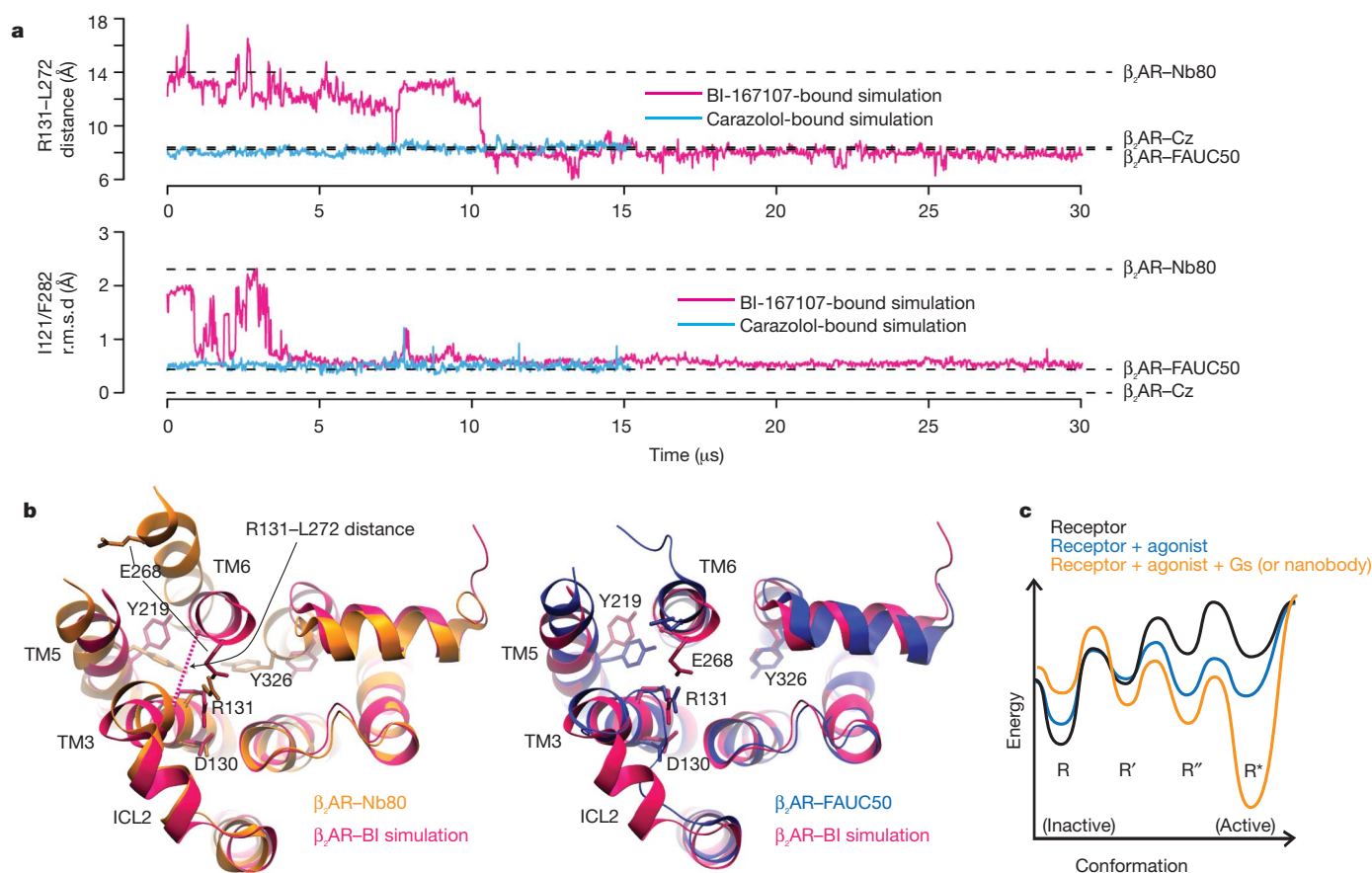


Figure 3 | Molecular dynamics simulations. **a**, An unbiased simulation initiated from the nanobody complex ($\beta_2\text{AR}$ -Nb80) structure, with the agonist BI-167107 bound but the nanobody removed (magenta), and a carazolol-bound simulation initiated from the inactive structure ($\beta_2\text{AR}$ -Cz) (blue). Top, distance between C_α atoms of Arg 131^{3,50} and Leu 272^{6,34}; and bottom, r.m.s.d. (root mean squared deviation) from the inactive structure of non-symmetric non-hydrogen atoms in residues Ile 121^{3,40} and Phe 282^{6,44}. Dashed lines indicate corresponding quantities from crystal structures. **b**, Cytoplasmic view

of the simulated agonist-bound receptor after 30 μs , compared to the $\beta_2\text{AR}$ -Nb80 (left) and $\beta_2\text{AR}$ -FAUC50 (right) structures. The conformations of intracellular loop 2, Tyr 219^{5,58} and Glu 268^{6,30} shown for the agonist-bound simulation differ from $\beta_2\text{AR}$ -FAUC50, but have been observed in inactive $\beta_2\text{AR}$ simulations¹⁹ and in other inactive-state GPCR structures¹. **c**, Proposed energy landscape model, in which both an agonist and a cytoplasmic binding partner are required to stabilize the fully active receptor conformation (R^*) over intermediate (R' and R'') and inactive (R) states.

conformation. Our crystallographic studies and MD simulations are compatible with fluorescence lifetime experiments on the purified receptor, demonstrating that saturating concentrations of a full agonist are incapable of pushing the $\beta_2\text{AR}$ conformational equilibrium towards a homogenous active state²³. These results indicate that binding energy from a G protein or nanobody interaction is required to stabilize conformational changes such as those observed in the vicinity of Ile 121^{3,40} and Phe 282^{6,44} in the active state (Fig. 2c). Alternative biophysical approaches such as NMR^{24,25} and further molecular dynamics simulations will be crucial to understand these transitions and identify potential pathways and intermediates that are compatible with these structures.

METHODS SUMMARY

Synthetic chemistry methodology to generate FAUC50 and FAUC72 is described in Supplementary Information. Protocols for radioligand binding and Gs activation assays to characterize ligands are described in Methods.

Crystallization. FAUC50- $\beta_2\text{AR}^{\text{H93C}}$ T4L was expressed, purified and crystallized as described in Methods. The receptor was crystallized in a cholesterol-doped (10%) monolein cubic phase overlaid with precipitant in glass sandwich plates. Optimized precipitant consisted of 24–27% (v/v) PEG 400, 200 mM Li_2SO_4 , 4% (v/v) DMSO, 3.5% (v/v) 1,4-butanediol, 100 mM MES pH 6.7. After 3–5 days of growth, crystals were harvested after adding an excess of precipitant solution for cryoprotection, and flash-frozen in liquid nitrogen.

Data collection, structure solution and refinement. Diffraction data were collected at beamline 23-ID (GM/CA-CAT) of the Advanced Photon Source, using a 10 μm diameter collimated microbeam. Oscillation data were measured in 1.0° frames with 10 s or 25 s exposures using 2 \times or 5 \times attenuated beam, respectively. The complete data set consisted of 106° of data from 19 crystals. The structure of FAUC50- $\beta_2\text{AR}^{\text{H93C}}$ T4L was solved by molecular replacement, and refined by group B factor and TLS refinement. Full details are provided in Methods.

Molecular dynamics. All-atom classical molecular dynamics simulations with explicitly represented lipids and water were performed using the CHARMM force field²⁶ on Anton²⁷, a special-purpose computer that accelerates such simulations by orders of magnitude; details are provided in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 August; accepted 11 November 2010.

- Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
- Caffrey, M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* **38**, 29–51 (2009).
- Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* doi:10.1038/nature09648 (this issue).
- Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).
- Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).

7. Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
8. Deupi, X. & Kobilka, B. K. Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology (Bethesda)* **25**, 293–303 (2010).
9. Dohlmann, H. G., Caron, M. G., Strader, C. D., Amlaiki, N. & Lefkowitz, R. J. Identification and sequence of a binding site peptide of the β_2 adrenergic receptor. *Biochemistry* **27**, 1813–1817 (1988).
10. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein coupled receptors. *Meth. Neurosci.* **25**, 366–428 (1995).
11. Buck, E. & Wells, J. A. Disulfide trapping to localize small-molecule agonists and antagonists for a G protein-coupled receptor. *Proc. Natl Acad. Sci. USA* **102**, 2719–2724 (2005).
12. Kolb, H. C., VanNieuwenhze, M. S. & Sharpless, K. B. Catalytic asymmetric dihydroxylation. *Chem. Rev.* **94**, 2483–2547 (1994).
13. Martinelli, M. J. *et al.* Catalytic regioselective sulfonation of α -chelatable alcohols: scope and mechanistic insight. *J. Am. Chem. Soc.* **124**, 3578–3585 (2002).
14. Whorton, M. R. *et al.* A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
15. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
16. Cherezov, V., Peddi, A., Muthusubramanian, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D* **60**, 1795–1807 (2004).
17. Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into β_2 -adrenergic receptor function. *Science* **318**, 1266–1273 (2007).
18. Tota, M. R., Candelore, M. R., Dixon, R. A. & Strader, C. D. Biophysical and genetic analysis of the ligand-binding site of the β -adrenoceptor. *Trends Pharmacol. Sci.* **12**, 4–6 (1991).
19. Dror, R. O. *et al.* Identification of two distinct inactive conformations of the β_2 -adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl Acad. Sci. USA* **106**, 4689–4694 (2009).
20. Vilardaga, J. P., Bunemann, M., Krasel, C., Castro, M. & Lohse, M. J. Measurement of the millisecond activation switch of G protein-coupled receptors in living cells. *Nature Biotechnol.* **21**, 807–812 (2003).
21. Vanni, S., Neri, M., Tavernelli, I. & Rothlisberger, U. A conserved protonation-induced switch can trigger “ionic-lock” formation in adrenergic receptors. *J. Mol. Biol.* **397**, 1339–1349 (2010).
22. Ghanouni, P. *et al.* The effect of pH on β_2 adrenoceptor function. Evidence for protonation-dependent activation. *J. Biol. Chem.* **275**, 3121–3127 (2000).
23. Ghanouni, P. *et al.* Functionally different agonists induce distinct conformations in the G protein coupling domain of the β_2 adrenergic receptor. *J. Biol. Chem.* **276**, 24433–24436 (2001).
24. Ahuja, S. *et al.* Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation. *Nature Struct. Mol. Biol.* **16**, 168–175 (2009).
25. Bokoch, M. P. *et al.* Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463**, 108–112 (2010).
26. MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
27. Shaw, D. E. *et al.* in *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis* (ACM Press, 2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from National Institutes of Health Grants NS028471 and GM083118 (B.K.K.), GM56169 (W.I.W.), P01 GM75913 (S.H.G.), GM75915 and P50GM073210 (M.C.), and P60DK-20572 (R.K.S.), the Mathers Foundation (B.K.K. and W.I.W.), the Lundbeck Foundation (Junior Group Leader Fellowship, S.G.F.R.), Science Foundation Ireland (07/IN.1/B1836) and FP7 COST Action CM0902 (M.C.), the Bavaria California Technology Center (P.G.), and the University of Michigan Biomedical Sciences Scholars Program (R.K.S.). We thank Stefan Löber, Harald Hübner, Albert Pan and Paul Maragakis for discussion and suggestions. We thank Foon Sun Thian for help with insect cell expression.

Author Contributions D.M.R. designed project and agonists, did binding assays to characterize agonists, developed purification, optimized crystallization conditions, optimized construct, purified protein, grew and harvested crystals, collected data, solved structure, refined structure, wrote manuscript. C.Z. performed G protein activation assays for covalent agonist, prepared recombinant baculovirus, performed large-scale expression of recombinant β_2 AR in insect cells, purified protein, grew crystals. J.L. helped to optimize crystallization conditions, grew and harvested crystals, collected data. D.A. was involved in LCP optimization, harvested crystals, collected data. R.H. synthesized agonists. S.G.F.R. identified the use of MNG3 detergent for β_2 AR stabilization and assisted with manuscript preparation. H.J.C. assisted with data processing and refinement. R.K.S. and B.D. provided ApoA1 and Gs protein for functional characterization of the covalent ligand. P.S.C. and S.G. provided MNG3 detergent for stabilization of purified β_2 AR. D.H.A. and R.O.D. performed and analyzed MD simulations; R.O.D. and D.E.S. oversaw MD simulations and analysis. W.I.W. assisted with data processing and refinement and with manuscript preparation. M.C. helped to guide the LCP crystallization efforts at Stanford and in Ireland, and oversaw automated lipidic cubic phase crystallography screens. P.G. designed the strategy for the synthesis of the covalent agonist. B.K. was responsible for the overall project strategy and management, oversaw manuscript preparation, and assisted with synchrotron data collection.

Author Information Coordinates and structure factors for β_2 AR-FAUC50 are deposited in the Protein Data Bank (accession code 3PDS). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.K.K. (kobilka@stanford.edu), M.C. (martin.caffrey@tcd.ie) or P.G. (peter.gmeiner@medchem.uni-erlangen.de).

METHODS

G protein activation assay with ICI-118,551 reversal. Wild-type β_2 AR or β_2 AR^{H93C} were purified in unliganded form²⁸. Samples of these receptors were reconstituted into rHDL (recombinant high density lipoprotein) particles as described¹⁴. For GTP γ S binding assay, receptor-rHDL particles were pre-incubated with 5 μ M FAUC50 or 5 μ M FAUC72 for 4 h at 4 °C. Samples were diluted 20-fold into binding buffer and split. Half of the samples were used in GTP γ S binding assays without ICI-118,551 competition, while the other half was incubated with 6 μ M ICI-118,551 at room temperature for 1 h with shaking. Control samples of receptor-rHDL particles with no ligand, 10 μ M isoproterenol, or 10 μ M ICI-118,551 were also prepared. Purified Gs heterotrimer²⁹ was added to each sample and incubated for 10 min at 23 °C. The final concentrations of reconstituted receptor and Gs were 100 nM and 600 nM, respectively. GTP γ S binding reactions were initiated by the addition of 0.4 nM [³⁵S]GTP γ S. Free [³⁵S]GTP γ S was removed by rapid filtration of the particles using glass fibre filters. Filter-bound radioactivity was determined by liquid scintillation counting using a Beckman LS6000 scintillation counter. Data shown in Fig. 1c are from three independent experiments each performed in triplicate.

Purification of β_2 AR^{H93C}T4L. The construct used for crystallography consisted of the β_2 AR-T4L fusion protein¹⁷ with the following modifications: a TEV (tobacco etch virus) protease site was inserted after residue 23 of the human β_2 AR sequence; Histidine 93 was mutated to cysteine; the construct was terminated by a 6 \times His tag after residue 348 of human β_2 AR. This construct was cloned into the baculovirus transfer vector pVL1392, and the resulting vector was used to make a high-titre baculovirus stock using the Bestbac system (Expression Systems). The β_2 AR^{H93C}T4L receptor was expressed in Sf9 insect cell cultures infected with this baculovirus, and Sf9 membranes were solubilized as described¹⁷. Purification was achieved using M1 Flag affinity chromatography (Sigma), functional alprenolol-Sepharose chromatography³⁰, and a second M1 chromatography concentrating step. While the receptor was still bound to the second M1 Flag column, bound alprenolol was washed out and replaced with the covalent ligand using 3 h of washing with 15 column volumes of buffer containing 30 μ M FAUC50 (tenfold total molar excess over receptor). Likewise, the dodecylmaltoside detergent used in all previous steps was exchanged for 0.1% (w/v) MNG-3 amphiphile³¹. Covalent agonist-bound and detergent-exchanged β_2 AR^{H93C}T4L was eluted in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.1% (w/v) MNG-3, 30 μ M FAUC50. In contrast to previously published protocols, receptor was not alkylated before alprenolol-Sepharose chromatography. Instead, receptor was alkylated after agonist exchange by treatment with 2 mM iodoacetamide for 30 min at 4 °C. Incubation of the eluate with AcTEV protease (Invitrogen) succeeded in removing the N terminus of the receptor (23 amino acids plus TEV site, leaving a glycine scar preceding residue 24), as verified by SDS-PAGE analysis. The sample at this stage was further purified and concentrated using Ni chelating chromatography, taking advantage of the C-terminal 6 \times His tag. Receptor was bound to a 0.7-ml column Ni-Sepharose column, and eluted in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.1% (w/v) MNG-3 and 30 μ M FAUC50, 200 mM imidazole. Finally, the receptor was concentrated to 50 mg ml⁻¹ with a 100 kDa cutoff Vivaspinn concentrator (Vivascience).

Crystallization. Purified FAUC50- β_2 AR^{H93C}T4L was crystallized using the *in meso* method². Cubic phase reconstitution consisted of two parts 50 mg ml⁻¹ receptor and three parts molten lipid mixture (10:1 monoolein:cholesterol by mass, lipids purchased from Sigma). Note that cholesterol is required for successful crystallization, and one cholesterol molecule is included in the refined model. Aqueous and lipid components were combined at room temperature using a syringe mixing apparatus³². Crystallization experiments in glass sandwich plates, set up by either by hand or using an *in meso* robot¹⁶, consisted of 30–50 nl cubic phase overlaid with 800 nl precipitant. Precipitant conditions producing diffraction quality crystals were identified by screening around previous conditions¹⁷, and testing additives and alternative buffers. Final optimized conditions consisted of 24–27% (v/v) PEG 400, 200 mM Li₂SO₄, 4% (v/v) DMSO, 3.5% (v/v) 1,4-butanediol, 100 mM MES pH 6.7. Crystals grew at 20 °C to a maximum size of 50 \times 15 \times 5 μ m³ within 3 to 5 days (see Supplementary Fig. 3, Supplementary Materials). For harvesting and cryocooling, exposed crystallization drops were overlaid with an excess of precipitant solution, and cryoloops (MiTeGen) containing single crystals were flash-frozen in liquid nitrogen.

Data collection and processing. Diffraction data were collected at beamline 23-ID of the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) of the Advanced Photon Source, Argonne, Illinois, USA. All data was acquired using a 10- μ m diameter collimated microbeam. Attenuated 1.0° rotation images were used to locate and centre crystals within the opaque mesophase in each cryoloop. Oscillation data were measured in 1.0° frames with 10 s or 25 s exposures using 2 \times or 5 \times attenuated beam, respectively. Significant radiation damage caused decay in the signal that prevented merging more than the first 5–10° of oscillation

data from each crystal. A total of 106° of data from 19 crystals were integrated with Mosflm³³ and merged with Scala³⁴.

Structure solution and refinement. Molecular replacement to obtain initial phases was performed with the program Phaser³⁵. The separated structures of the receptor and T4L components of the high-resolution carazolol-bound β_2 AR structure (PDB ID 2RH1 with all non-protein atoms removed), were used as search models. The model was refined in Phenix³⁶ and Buster³⁷, using group B factor refinement (one B factor per residue) followed by TLS refinement (using two TLS groups, one for the receptor and one for T4L). Refinement statistics are given in Supplementary Table 1 of Supplementary Information. The crystallographic data was strongly anisotropic, as seen in the anisotropic B factor corrections; however, the electron density was clear enough for the placement of side chains. Although there was clear electron density present for the agonist (see omit map in Supplementary Fig. 4, Supplementary Information), we did observe a small discontinuity at the end of the linker connecting to Cys 93. This discontinuity could arise from flexibility of the polymethylene component of the linker, as well as potential radiation damage³⁸ localized to the disulphide bond. As a control, we carried out refinement in which the ligand was modelled as a non-covalent species without a disulphide connection to the receptor. No significant differences in final model R/R_{free} or $2Fo - Fc$ and $Fo - Fc$ electron density maps were observed.

Molecular dynamics simulations. In all simulations, β_2 AR was embedded in a hydrated lipid bilayer, with all atoms, including those in the lipids and water, represented explicitly. The BI-167107-bound β_2 AR simulations were initiated with the receptor and ligand in the conformation of the β_2 AR-Nb80 crystal structure (companion paper³), with the nanobody (Nb80) removed. The carazolol-bound β_2 AR simulations were initiated with the receptor and ligand in the conformation of the high-resolution carazolol-bound crystal structure¹⁷. Both crystal structures were determined using a β_2 AR-T4L fusion protein, in which intracellular loop 3 (ICL3) of the receptor was replaced by T4 lysozyme (although the T4L was not resolved in the BI-167107-bound nanobody complex structure). We omitted the T4L in all simulations. Experimentally, removal of the bulk of ICL3 by partial tryptic digest does not seem to affect receptor function³⁹.

Production simulations were performed on Anton²⁷, a special-purpose computer designed to accelerate standard molecular dynamics simulations by orders of magnitude relative to the previous state of the art. Prior to production simulation, systems were equilibrated using Desmond⁴⁰ on a commodity cluster, according to the protocol described below.

Hydrogens were added to the crystal structures of carazolol-bound β_2 AR (β_2 AR-Cz; PDB ID 2RH1) and β_2 AR-Nb80/BI-167107 (companion paper³) using Maestro (Schrödinger) as described previously¹⁹. T4L and Nb80 were deleted, and chain termini were capped with neutral groups (acetyl and methylamino).

All titratable residues other than Glu 122^{3,41}, Asp 79^{2,50}, and Asp 130^{3,49} were left in the dominant protonation state at pH 7.0. Glu 122^{3,41} was protonated in all simulations. It faces the lipid bilayer and is thus likely protonated⁴¹; in addition, a similarly positioned residue in rhodopsin (Glu 122^{3,37}) has been found to be protonated during the entire photocycle²².

We performed both BI-167107-bound and carazolol-bound simulations using two different sets of protonation states for Asp 79^{2,50} and Asp 130^{3,49}. Previous studies have suggested that Asp 130^{3,49} is protonated upon activation²², and FTIR (Fourier transform infrared) spectroscopy data has shown that this is the case for the corresponding residue of rhodopsin, Glu 134^{3,49} (ref. 42). Asp 79^{2,50} is homologous to Asp 83^{2,50} of rhodopsin, which has been found by FTIR spectroscopy to remain protonated during the entire photocycle⁴¹. On the other hand, neutralization of Asp 79^{2,50} in β_2 AR by mutation to asparagine uncouples agonist binding from G protein activation⁴³, and a recent study suggested that Asp 79 may be deprotonated upon activation²¹. Thus, we repeated each simulation with two sets of protonation states: first with Asp 79 protonated and Asp 130 deprotonated ('Ash 79/Asp 130', representing potential protonation states in the inactive receptor), and then with Asp 79 deprotonated and Asp 130 protonated ('Asp 79/Ash 130', representing possible protonation states in the active receptor). Results shown in Supplementary Figs 3, 5 and 6 are for the Ash 79/Asp 130 simulations, whereas results in Supplementary Fig. 7 are for the Asp 79/Ash 130 simulations.

β_2 AR residues that were truncated or not resolved in the crystal structures were omitted from the simulations. In particular, N-terminal residues 1–28, C-terminal residues 343–413, and ICL3 residues 231–262 were omitted from the carazolol-bound system, whereas N-terminal residues 1–22, C-terminal residues 345–413 and ICL3 residues 228–265 were omitted from the BI-167107-bound system. Both simulations included a glutamate at position 187, reflecting an Asn 187 Glu mutation made in both crystallization constructs to eliminate a glycosylation site.

We prepared the carazolol-bound Ash 79/Asp 130 β_2 AR system following our previously described protocol¹⁹, and the other systems according to an updated protocol (details below). We do not expect the differences in the simulation setup

protocols to affect the quantities of interest in this study. As a control, however, we performed an additional 5 μ s simulation of a carazolol-bound Ash 79/Asp 130 β_2 AR system prepared according to the new protocol and obtained results (not shown) similar to those of the 15 μ s carazolol-bound Ash 79/Asp 130 β_2 AR simulation prepared according to the old protocol.

Prepared protein structures were inserted into an equilibrated phospholipid bilayer solvated with 0.15 M NaCl as described previously¹⁹. Carazolol-bound Ash 79/Asp 130 β_2 AR was simulated in a POPE (palmitoyl oleoyl phosphatidyl ethanolamine) bilayer system initially measuring $85 \times 73 \times 88$ Å, containing 170 lipid molecules, 8,798 water molecules, 19 sodium ions, and 24 chloride ions, for a total of 52,396 atoms. Carazolol-bound Asp 79/Ash 130 β_2 AR was simulated in a POPC (palmitoyl oleoyl phosphatidyl choline) bilayer system initially measuring $83 \times 83 \times 87$ Å, containing 160 lipid molecules, 11,313 water molecules, 30 sodium ions, and 35 chloride ions, for a total of 60,153 atoms. (The 5 μ s carazolol-bound Ash 79/Asp 130 control simulation prepared according to the new protocol also used a POPC bilayer.) BI-167107-bound β_2 AR (both Ash 79/Asp 130 and Asp 79/Ash 130) was simulated in a POPC bilayer system initially measuring $79 \times 79 \times 87$ Å, containing 138 lipid molecules, 10,116 water molecules, 26 sodium ions, and 31 chloride ions, for a total of 53,603 atoms.

All systems were equilibrated in the NPT ensemble at 310 K and 1 bar using the Berendsen coupling scheme with $5 \text{ kcal mol}^{-1} \text{ Å}^{-2}$ harmonic position restraints applied to all non-hydrogen atoms of the protein; these restraints were tapered off linearly over 5 ns. Unrestrained systems were then simulated for an additional 5 ns to equilibrate the aspect ratio of the simulation box further. During the equilibration process, van der Waals and short-range electrostatic interactions were cut off at 9 Å for the carazolol-bound Ash 79/Asp 130 β_2 AR system and at 12 Å for the other systems. Long-range electrostatic interactions were computed using the particle mesh Ewald method⁴⁴, with a $64 \times 64 \times 64$ grid and $\sigma = 2.26$ Å for the carazolol-bound Ash 79/Asp 130 β_2 AR system, and with a $32 \times 32 \times 32$ grid and $\sigma = 3.23$ Å for the other systems; fifth-order B-splines were used for interpolation in both cases. All bond lengths to hydrogen atoms were constrained using M-SHAKE⁴⁵. A RESPA integrator⁴⁶ was used with a time step of 2 fs, and long-range electrostatics were computed every 6 fs.

Production simulations on Anton were initiated from the final snapshot of each corresponding equilibration run on Desmond, using the same integration schemes. Van der Waals and short-range electrostatic interactions were cut off at 9 Å for the carazolol-bound Ash 79/Asp 130 β_2 AR system and at 13.5 Å for the other systems. Long-range electrostatics were computed using the *k*-space Gaussian split Ewald method⁴⁷, with a $64 \times 64 \times 64$ grid, $\sigma = 2.01$ Å, and $\sigma_s = 1.41$ Å for the carazolol-bound Ash 79/Asp 130 β_2 AR system, and with a $32 \times 32 \times 32$ grid, $\sigma = 3.33$ Å, and $\sigma_s = 2.35$ Å for the other systems.

Force field parameters: the CHARMM27 (ref. 26) parameter set (with CMAP terms⁴⁸) and the CHARMM TIP3P⁴⁹ water model were used for all protein molecules, POPE lipid molecules, water molecules, and salt ions. A modified CHARMM lipid force field⁵⁰, which became available after we performed the 15- μ s simulation of carazolol-bound β_2 AR in POPE, was used for POPC lipids. Force-field parameters for carazolol and palmitoyl-cysteine were designed previously¹⁹. Force-field parameters for BI-167107 were transferred from previously parameterized model compounds: parameters for the hydroxyethylamine 'tail' group were transferred from the alkylamine parameters we previously designed for carazolol¹⁹, and parameters for the benzoxazine 'head' group were transferred from the model compounds anisole and *p*-phenol acetamide from the CHARMM General Force Field⁵¹. Full parameter sets are available upon request.

Analysis protocols: trajectory snapshots, each containing a record of all atom positions at a particular instant in time, were saved every 180 ps during production simulation. Distance and r.m.s.d. measurements were computed using the HiMach parallel analysis framework⁵².

The distance and r.m.s.d. measurements shown in Fig. 3a, Supplementary Fig. 5 and Supplementary Fig. 7 are smoothed; the smoothed time series were computed from the original time series by a weighted running average, using a filter kernel of

half-width 12.51 ns whose shape corresponded to that of the cosine function from $-\pi/2$ to $\pi/2$.

VMD⁵³ was used to visualize trajectories and to produce the molecular renderings of Fig. 3b.

28. Yao, X. *et al.* Coupling ligand structure to specific conformational switches in the β_2 -adrenoceptor. *Nature Chem. Biol.* **2**, 417–422 (2006).
29. Kozasa, T. & Gilman, A. G. Purification of recombinant G proteins from Sf9 cells by hexahistidine tagging of associated subunits. Characterization of α_{12} and inhibition of adenylyl cyclase by α_2 . *J. Biol. Chem.* **270**, 1734–1741 (1995).
30. Kobilka, B. K. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. *Anal. Biochem.* **231**, 269–271 (1995).
31. Chae, P. S. *et al.* Maltose–neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature Methods* **7**, 1003–1008 (2010).
32. Cheng, A., Hummel, B., Qiu, H. & Caffrey, M. A simple mechanical mixer for small viscous liquid-containing samples. *Chem. Phys. Lipids* **95**, 11–21 (1998).
33. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
34. Collaborative Computational Project, number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
36. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D* **61**, 850–855 (2005).
37. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
38. Weik, M. *et al.* Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl Acad. Sci. USA* **97**, 623–628 (2000).
39. Rubenstein, R. C. *et al.* The hydrophobic tryptic core of the β -adrenergic receptor retains G_s regulatory activity in response to agonists and thiols. *J. Biol. Chem.* **262**, 16655–16662 (1987).
40. Bowers, K. J. *et al.* in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing* (ACM Press, 2006).
41. Fahmy, K. *et al.* Protonation states of membrane-embedded carboxylic acid groups in rhodopsin and metarhodopsin II: a Fourier-transform infrared spectroscopy study of site-directed mutants. *Proc. Natl Acad. Sci. USA* **90**, 10206–10210 (1993).
42. Vogel, R. *et al.* Functional role of the “ionic lock”—an interhelical hydrogen-bond network in family A heptahelical receptors. *J. Mol. Biol.* **380**, 648–655 (2008).
43. Liapakis, G., Chan, W. C., Papadokostaki, M. & Javitch, J. A. Synergistic contributions of the functional groups of epinephrine to its affinity and efficacy at the β_2 adrenergic receptor. *Mol. Pharmacol.* **65**, 1181–1190 (2004).
44. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N -log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
45. Krätter, V., van Gunsteren, W. F. & Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **22**, 501–508 (2001).
46. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990 (1992).
47. Shan, Y., Klepeis, J. L., Eastwood, M. P., Dror, R. O. & Shaw, D. E. Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* **122**, 054101 (2005).
48. Mackerell, A. D. Jr, Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
49. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063 (1994).
50. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
51. Vanommeslaeghe, K. *et al.* CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
52. Tu, T. *et al.* in *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing* (ACM Press, 2008).
53. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

The structural basis for agonist and partial agonist action on a β_1 -adrenergic receptor

Tony Warne¹, Rouslan Moukhametzianov¹, Jillian G. Baker², Rony Nehmé¹, Patricia C. Edwards¹, Andrew G. W. Leslie¹, Gebhard F. X. Schertler^{1†} & Christopher G. Tate¹

β -adrenergic receptors (β ARs) are G-protein-coupled receptors (GPCRs) that activate intracellular G proteins upon binding catecholamine agonist ligands such as adrenaline and noradrenaline^{1,2}. Synthetic ligands have been developed that either activate or inhibit β ARs for the treatment of asthma, hypertension or cardiac dysfunction. These ligands are classified as either full agonists, partial agonists or antagonists, depending on whether the cellular response is similar to that of the native ligand, reduced or inhibited, respectively. However, the structural basis for these different ligand efficacies is unknown. Here we present four crystal structures of the thermostabilized turkey (*Meleagris gallopavo*) β_1 -adrenergic receptor (β_1 AR-m23) bound to the full agonists carmoterol and isoprenaline and the partial agonists salbutamol and dobutamine. In each case, agonist binding induces a 1 Å contraction of the catecholamine-binding pocket relative to the antagonist bound receptor. Full agonists can form hydrogen bonds with two conserved serine residues in transmembrane helix 5 (Ser^{5.42} and Ser^{5.46}), but partial agonists only interact with Ser^{5.42} (superscripts refer to Ballesteros–Weinstein numbering³). The structures provide an understanding of the pharmacological differences between different ligand classes, illuminating how GPCRs function and providing a solid foundation for the structure-based design of novel ligands with predictable efficacies.

Determining how agonists and antagonists bind to the β receptors has been the goal of research for more than 20 years^{4–11}. Although the structures of the homologous β_1 and β_2 receptors^{12–15} show how some antagonists bind to receptors in the inactive state¹⁶, structures with agonists bound are required to understand subsequent structural transitions involved in activation. GPCRs exist in an equilibrium between an inactive state (R) and an activated state (R*) that can couple and activate G proteins¹⁷. The binding of a full agonist, such as adrenaline or noradrenaline, is thought to increase the probability of the receptor converting to R*, with a conformation similar to that of opsin^{18,19}. In the absence of any ligand, the β ARs exhibit a low level of constitutive activity, indicating that there is always a small proportion of the receptor in the activated state, with the β_2 AR showing a fivefold higher level of basal activity than the β_1 AR²⁰. Basal activity of β_2 AR is important physiologically, as shown by the T164I^{4.56} human polymorphism that reduces the basal activity of β_2 AR to levels similar to β_1 AR²¹ and whose expression has been associated with heart disease²².

As a first step towards understanding how agonists activate receptors, we have determined the structures of β_1 AR bound to four different agonists. Native turkey β_1 AR is unstable in detergent²³, so crystallization and structure determination relied on using a thermostabilized construct (β_1 AR-m23) that contained six point mutations, which dramatically improved its thermostability²⁴. In addition, the thermostabilizing mutations altered the equilibrium between R and R*, so that the receptor was preferentially in the R state²⁴. However, it could still couple to G proteins after activation by agonists¹³ (Supplementary Fig. 1 and Supplementary Tables 1–3), although the activation energy barrier is predicted to be considerably higher than for the wild-type receptor²⁵.

Here we report structures of β_1 AR-m23 (see Methods) bound to R-isoprenaline (2.85 Å resolution), R,R-carmoterol (2.6 Å resolution), R-salbutamol (3.05 Å resolution) and R-dobutamine (two independent structures at 2.6 Å and 2.5 Å resolution) (Supplementary Table 5). The overall structures of β_1 AR-m23 bound to the agonists are very similar to the structure with the bound antagonist cyanopindolol¹³, as expected for a receptor mutant stabilized preferentially in the R state. None of the structures show the outward movement of the cytoplasmic end of transmembrane helix H6 by 5–6 Å that is observed during light activation of rhodopsin^{18,19,26}. This indicates that the structures represent an inactive, non-signalling state of the receptor formed on initial agonist binding.

All four agonists bind in the catecholamine pocket in a virtually identical fashion (Fig. 1). The secondary amine and β -hydroxyl groups shared by all the agonists (except for dobutamine, which lacks the β -hydroxyl; see Supplementary Fig. 4) form potential hydrogen bonds with Asp 121^{3.32} and Asn 329^{7.39}, whereas the hydrogen bond donor/acceptor group equivalent to the catecholamine *meta*-hydroxyl (*m*-OH) generally forms a hydrogen bond with Asn 310^{6.55}. In addition, all the agonists can form a hydrogen bond with Ser 211^{5.42}, as seen for cyanopindolol¹³, and they also induce the rotamer conformation change of Ser 212^{5.43} so that it makes a hydrogen bond with Asn 310^{6.55}. The major difference between the binding of full agonists compared to the partial agonists is that only full agonists make a hydrogen bond to the side chain of Ser 215^{5.46} as a result of a change in side chain rotamer. All of these amino acid residues involved in the binding of the catecholamine head-groups to β_1 AR are fully conserved in both β_1 and β_2 receptors (Fig. 2). Furthermore, the role of many of these amino acid residues in ligand binding is supported by extensive mutagenesis studies on β_2 AR that were performed before the first β_2 AR structure was determined²⁷. Thus it was predicted that Asp 113^{3.32}, Ser 203^{5.42}, Ser 207^{5.46}, Asn 293^{6.55} and Asn 312^{7.39} in β_2 AR were all involved in agonist binding^{4,5,7–9} (Fig. 3). Inspection of the region outside the catecholamine binding pocket in the structures with bound dobutamine and carmoterol allows the identification of non-conserved residues that interact with these ligands (Fig. 2 and Supplementary Fig. 7), which may contribute to the subtype specificity of these ligands^{10,28}.

There are three significant differences in the β_1 AR catecholamine binding pocket when full agonists are bound compared to when an antagonist is bound, namely the rotamer conformation changes of side chains Ser 212^{5.43} and Ser 215^{5.46} (Fig. 3) and the contraction of the catecholamine binding pocket by ~1 Å, as measured between the C α atoms of Asn 329^{7.39} and Ser 211^{5.42} (Fig. 4). So why should these small changes increase the probability of R* formation? Agonist binding has not changed the conformation of transmembrane helix H5 below Ser 215^{5.46}, although significant changes in this region are predicted once the receptor has reached the fully activated state^{18,19}. The only effect that the agonist-induced rotamer conformation change of Ser 215^{5.46} appears to have is to break the van der Waals interaction between Val 172^{4.56} and Ser 215^{5.46}, thus reducing the number of interactions

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. ²Institute of Cell Signalling, C Floor Medical School, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK. [†]Present address: Paul Scherrer Institut, Laboratory of Biomolecular Research, BMR, OFLC 109, CH-5232 Villigen PSI, Switzerland.

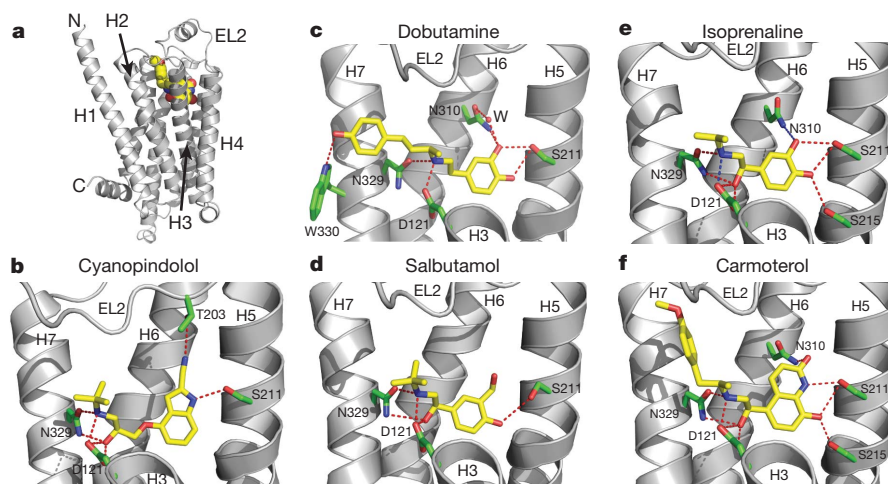


Figure 1 | Structure of the β_1 -adrenergic receptor bound to agonists.

a, Structure of β_1 AR shown in cartoon representation with the intracellular side at the bottom of the figure. The ligand carmoterol is shown as a space filling model (C, yellow; O, red; N, blue). The amino terminus (N), carboxy terminus (C), extracellular loop 2 (EL2), and transmembrane helices 1–4 (H1–H4) are labelled. **b–f**, The same orientation of receptor is shown in panels **b**, the antagonist cyanopindolol; **c**, **d**, the partial agonists dobutamine and salbutamol; **e**, **f**, the full agonists isoprenaline and carmoterol. The colour scheme of the

ligand and labelling of the receptor is identical in all panels, with amino acid side chains that make hydrogen bonds to the ligands depicted (C, green; O, red; N, blue). For clarity, residues 171–196 and 94–119 have been removed in **b–f**, which correspond to the C-terminal region of H4 and EL2, and EL1 with the C-terminal region of H2 and N-terminal region of H3, respectively. All structures shown are of monomer B (Supplementary Fig. 2) and were generated using Pymol (DeLano Scientific). For a comparison of the positions of the ligands when bound to the receptor, see Supplementary Fig. 5.

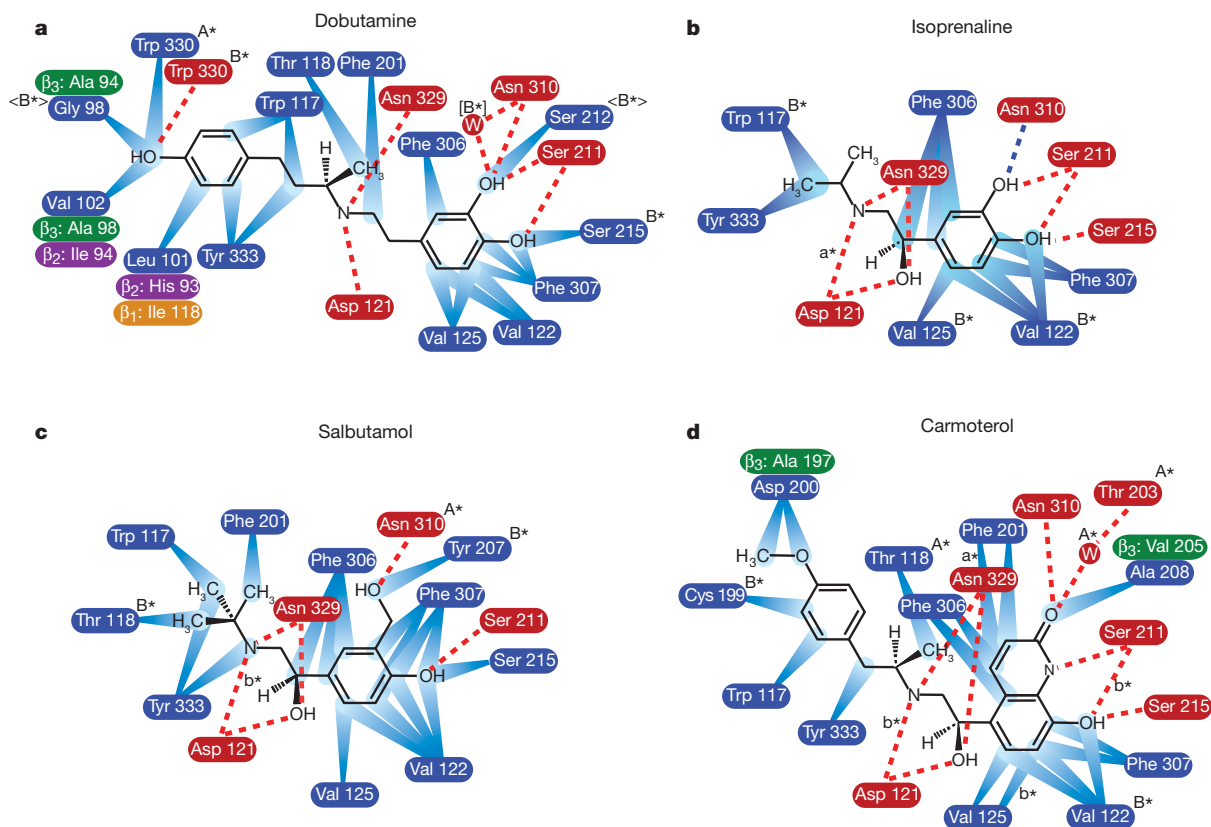


Figure 2 | Polar and non-polar interactions involved in agonist binding to β_1 -adrenergic receptor. **a–d**, Amino acid residues within 3.9 Å of the ligands are depicted, with residues highlighted in blue making van der Waals contacts (blue rays) and residues highlighted in red making potential hydrogen bonds with favourable geometry (red dashed lines) or hydrogen bonds with unfavourable geometry (blue dashed lines). Amino acid residues labelled with an asterisk make the indicated contact either in monomer A (A*) or in monomer B (B*) only; for dobutamine, some contacts, labelled <B*>, are

found only in monomer B of dob92, whereas another contact, labelled [B*], is found only in monomer B of dob102 (Supplementary Fig. 6 and also see Supplementary Table 6 for further details and for the Ballesteros–Weinstein numbering). If specific van der Waals interactions or polar interactions are found only in monomer A or B, then the interaction is labelled a* or b*, respectively. Where the amino acid residue differs between the turkey β_1 AR and the human β_1 AR, β_2 AR and β_3 AR, the equivalent residue is shown highlighted in orange, purple or green, respectively (see also Supplementary Table 7).

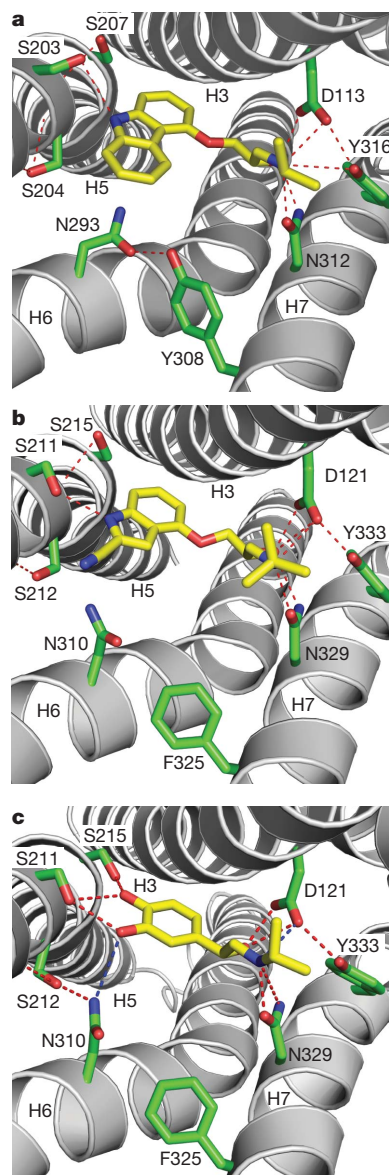


Figure 3 | Comparison of the ligand-binding pockets of the β_1 and β_2 adrenergic receptors. The ligand-binding pockets are shown as viewed from the extracellular surface with EL2 removed for clarity (same colour scheme as in Fig. 1). **a**, β_2 AR with the antagonist carazolol bound (PDB code 2RH1); **b**, β_1 AR with the antagonist cyanopindolol bound (PDB code 2VT4); **c**, β_1 AR with the agonist isoprenaline bound.

between H4 and H5. As there is only a minimal interface between transmembrane helices H4 and H5 in this region (Supplementary Table 8 and Supplementary Fig. 8), this loss of interaction may be significant in the activation process. In this regard, it is noteworthy that the naturally occurring polymorphism in β_2 AR at the H4–H5 interface, T164I^{4,56}, converts a polar residue to a hydrophobic residue as seen in β_1 AR (Val 172^{4,56}), which results in both reduced basal activity and reduced agonist stimulation²¹. This supports the hypothesis that the extent of interaction between H4 and H5 could affect the probability of a receptor transition into the activated state.

In contrast to the apparent weakening of helix–helix interactions by the agonist-induced rotamer conformation change of Ser 215^{5,46}, the agonist-induced rotamer conformation change of Ser 212^{5,43} probably results in the strengthening of interactions between H5 and H6. Upon agonist binding, Ser 212^{5,43} forms a hydrogen bond with Asn 310^{6,55} (Fig. 3) and, in addition, hydrogen bond interactions to Ser 211^{5,43} and Asn 310^{6,55} mediated by the ligand serve to bridge H5 and H6. The

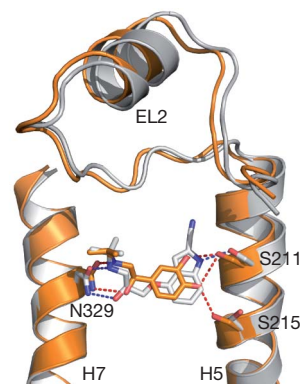


Figure 4 | Differences in the ligand-binding pocket between antagonist- and agonist-bound β_1 -adrenergic receptor. An alignment was performed (see Methods) between the structures of β_1 AR-m23 bound to either cyanopindolol (grey) or isoprenaline (orange) and the relative positions of the ligands and the transmembrane helices H5 and H7 are depicted. The 1 Å contraction of the ligand-binding pocket between H5 and H7 is clear.

combined effects of strengthening the H5–H6 interface and weakening the H4–H5 interface could facilitate the subsequent movements of H5 and H6, as observed in the activation of rhodopsin.

Stabilization of the contracted catecholamine binding pocket is probably the most important role of bound agonists in the activation process (Fig. 4). This probably requires strong hydrogen bonding interactions between the catechol (or equivalent) moiety and both H5 and H6, and strong interactions between the secondary amine and β -hydroxyl groups in the agonist and the amino acid side chains in helices H3 and H7. Reduction in the strength of these interactions is likely to reduce the efficacy of a ligand²⁹. Both salbutamol and dobutamine are partial agonists of β_1 AR-m23 (Supplementary Table 3) and human β_1 AR. In the case of salbutamol, there are only two predicted hydrogen bonds between the headgroup and H5/H6, compared to three–four potential hydrogen bonds for isoprenaline and carmoterol. Dobutamine lacks the β -hydroxyl group, which similarly reduces the number of potential hydrogen bonds to H3/H7 from three–four seen in the other agonists to only two. We propose that this weakening of agonist interactions with H5/H6 for salbutamol and H3/H7 for dobutamine is a major contributing factor in making these ligands partial agonists rather than full agonists.

The agonist-bound structures of β_1 AR indicate there are three major determinants that dictate the efficacy of any ligand: ligand-induced rotamer conformational changes of (1) Ser 212^{5,43} and (2) Ser 215^{5,46} and (3) stabilization of the contracted ligand-binding pocket. The full agonists studied here achieve all three. The partial agonists studied here do not alter the conformation of Ser 215^{5,46} and may be less successful than isoprenaline or carmoterol at stabilizing the contracted catecholamine binding pocket due to reduced numbers of hydrogen bonds between the ligand and the receptor. The antagonist cyanopindolol acts as a very weak partial agonist and none of the three agonist-induced changes are observed. In contrast to partial agonists, neutral antagonists or very weak partial agonists such as cyanopindolol may also have a reduced ability to contract the binding pocket owing to the greater distance between the secondary amine and the catechol moiety (or equivalent). For example, the number of atoms in the linker between the secondary amine and the headgroup of cyanopindolol is four, whereas the agonists in this study only contain two (Fig. 1 and Supplementary Fig. 4). A ligand with a sufficiently bulky headgroup that binds with high affinity and which actively prevents any spontaneous contraction of the binding pocket and/or Ser^{5,46} rotamer change, would be predicted to act as a full inverse agonist. This is indeed what is observed in the recently determined structure¹⁵ of β_2 AR bound to the inverse agonist ICI 118,551.

The significant structural similarities amongst GPCRs suggests that similar agonist-induced conformational changes to those we have

observed here may also be applicable to many other members of the GPCR superfamily, though undoubtedly there will be many subtle variations on this theme.

METHODS SUMMARY

Expression, purification and crystallization. The $\beta 44$ -m23 construct was expressed in insect cells, purified in the detergent Hega-10 and crystallized in the presence of cholesterol hemisuccinate (CHS), following previously established protocols³⁰. Crystals were grown by vapour diffusion, with the conditions shown in Supplementary Table 4.

Data collection, structure solution and refinement. Diffraction data were collected from a single cryo-cooled crystal (100 K) of each complex in multiple wedges at beamline ID23-2 at ESRF, Grenoble, France. The structures were solved by molecular replacement using the β_1 AR structure¹³ (PDB code 2VT4) as a model (see Methods). Data collection and refinement statistics are presented in Supplementary Table 5.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 July; accepted 9 December 2010.

- Evans, B. A. *et al.* Ligand-directed signalling at β -adrenoceptors. *Br. J. Pharmacol.* **159**, 1022–1038 (2010).
- Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Strader, C. D. *et al.* Conserved aspartic acid residues 79 and 113 of the β -adrenergic receptor have different roles in receptor function. *J. Biol. Chem.* **263**, 10267–10271 (1988).
- Sato, T., Kobayashi, H., Nagao, T. & Kurose, H. Ser²⁰³ as well as Ser²⁰⁴ and Ser²⁰⁷ in fifth transmembrane domain of the human β_2 -adrenoceptor contributes to agonist binding and receptor activation. *Br. J. Pharmacol.* **128**, 272–274 (1999).
- Liapakis, G. *et al.* The forgotten serine. A critical role for Ser-203^{5,42} in ligand binding to and activation of the β_2 -adrenergic receptor. *J. Biol. Chem.* **275**, 37779–37788 (2000).
- Strader, C. D. *et al.* Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor. *J. Biol. Chem.* **264**, 13572–13578 (1989).
- Wieland, K. *et al.* Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. *Proc. Natl Acad. Sci. USA* **93**, 9276–9281 (1996).
- Suryanarayana, S. & Kobilka, B. K. Amino acid substitutions at position 312 in the seventh hydrophobic segment of the beta 2-adrenergic receptor modify ligand-binding specificity. *Mol. Pharmacol.* **44**, 111–114 (1993).
- Kikkawa, H., Isogaya, M., Nagao, T. & Kurose, H. The role of the seventh transmembrane region in high affinity binding of a β_2 -selective agonist TA-2005. *Mol. Pharmacol.* **53**, 128–134 (1998).
- Isogaya, M. *et al.* Identification of a key amino acid of the β_2 -adrenergic receptor for high affinity binding of salmeterol. *Mol. Pharmacol.* **54**, 616–622 (1998).
- Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2 -adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
- Warne, T. *et al.* Structure of a β_1 -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
- Hanson, M. A. *et al.* A specific cholesterol binding site is established by the 2.8 Å structure of the human β_2 -adrenergic receptor. *Structure* **16**, 897–905 (2008).
- Wacker, D. *et al.* Conserved binding mode of human β_2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.* **132**, 11443–11445 (2010).
- Tate, C. G. & Schertler, G. F. Engineering G protein-coupled receptors to facilitate their structure determination. *Curr. Opin. Struct. Biol.* **19**, 386–395 (2009).
- Kobilka, B. K. & Deupi, X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **28**, 397–406 (2007).
- Park, J. H. *et al.* Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).
- Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
- Engelhardt, S., Grimmer, Y., Fan, G. H. & Lohse, M. J. Constitutive activity of the human β_1 -adrenergic receptor in β_1 -receptor transgenic mice. *Mol. Pharmacol.* **60**, 712–717 (2001).
- Green, S. A., Rathz, D. A., Schuster, A. J. & Liggett, S. B. The Ile164 β_2 -adrenoceptor polymorphism alters salmeterol exosite binding and conventional agonist coupling to G_s. *Eur. J. Pharmacol.* **411**, 141–147 (2001).
- Piscione, F. *et al.* Effects of Ile164 polymorphism of beta₂-adrenergic receptor gene on coronary artery disease. *J. Am. Coll. Cardiol.* **52**, 1381–1388 (2008).
- Serrano-Vega, M. J. & Tate, C. G. Transferability of thermostabilizing mutations between β -adrenergic receptors. *Mol. Membr. Biol.* **26**, 385–396 (2009).
- Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the β_1 -adrenergic receptor in a detergent-resistant form. *Proc. Natl Acad. Sci. USA* **105**, 877–882 (2008).
- Balaraman, G. S., Bhattacharya, S. & Nagarajan, V. Structural insights into conformational stability of wild-type and mutant β_1 -adrenergic receptor. *Biophys. J.* **99**, 568–577 (2010).
- Altenbach, C. *et al.* High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc. Natl Acad. Sci. USA* **105**, 7439–7444 (2008).
- Rasmussen, S. G. *et al.* Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
- Williams, R. S. & Bishop, T. Selectivity of dobutamine for adrenergic receptor subtypes: *in vitro* analysis by radioligand binding. *J. Clin. Invest.* **67**, 1703–1711 (1981).
- Katritch, V. *et al.* Analysis of full and partial agonists binding to β_2 -adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes. *J. Mol. Recognit.* **22**, 307–318 (2009).
- Warne, T., Serrano-Vega, M. J., Tate, C. G. & Schertler, G. F. Development and crystallization of a minimal thermostabilised G protein-coupled receptor. *Protein Expr. Purif.* **65**, 204–213 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. This work was supported by core funding from the MRC and the BBSRC grant (BB/G003653/1). Financial support for G.F.X.S. was also from a Human Frontier Science Project (HFSP) programme grant (RG/0052), a European Commission FP6 specific targeted research project (LSH-2003-1.1.0-1) and an ESRF long-term proposal. J.G.B. was funded by a Wellcome Trust Clinician Scientist Fellowship. We are grateful to P. Coli and A. Rizzi for the supply of (R,R)-carmoterol. F. Gorrec is thanked for his help with crystallisation robotics. We would also like to thank beamline staff at the European Synchrotron Radiation Facility, particularly D. Flot and A. Popov at ID23-2 and F. Marshall, M. Weir, M. Congreve and R. Henderson for comments on the manuscript.

Author Contributions T.W. devised and performed receptor expression, purification, crystallization, cryo-cooling of the crystals, data collection and initial data processing. P.C.E. helped with crystal cryo-cooling and data collection. J.G.B. performed the pharmacological analyses on receptor mutants in whole cells and R.N. performed the ligand binding studies on baculovirus-expressed receptors. R.M. and A.G.W.L. were involved in data processing and structure refinement. Manuscript preparation was performed by T.W., C.G.T., A.G.W.L. and G.F.X.S. The overall project management was by G.F.X.S. and C.G.T.

Author Information Coordinates and structure factors have been submitted to the PDB database under accession codes 2y00, 2y01, 2y02, 2y03 and 2y04 for $\beta 44$ -m23 bound either to dobutamine (dob92 and dob102), carmoterol, isoprenaline or salbutamol, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.G.T. (cgt@mrc-lmb.cam.ac.uk) or G.F.X.S. (gebhard.schertler@psi.ch).

METHODS

Expression, purification and crystallization. The turkey (*M. gallopavo*) β_1 AR construct, β_36 -m23, contains six thermostabilizing point mutations and truncations at the N terminus, inner loop 3 and C terminus³⁰. Here we used the β_44 -m23 construct, which differs from the previously published β_36 -m23 construct only by the presence of two previously deleted amino acid residues at the cytoplasmic end of helix 6 (H6), Thr 277 and Ser 278. Baculovirus expression and purification were all performed as described previously³⁰, but with the detergent exchanged to Hega-10 (0.35%) on the alprenolol affinity column. Purified receptor was competitively eluted from the alprenolol Sepharose column with 0.2 mM agonist ((R)-isoprenaline, (R,S)-salbutamol, (R,S)-dobutamine or (R,R)-carmoterol). The buffer was exchanged to 10 mM Tris-HCl, pH 7.7, 100 mM NaCl, 0.1 mM EDTA, 0.35% Hega-10 and 1.0 mM agonist during concentration to 15–20 mg ml⁻¹. Before crystallization, CHS and Hega-10 were added to 0.45–1.8 mg ml⁻¹ and 0.5–0.65%, respectively. Crystals were grown at 4 °C in 200 nl sitting drops and cryo-protected by soaking in either PEG 400 or PEG 600 for ~5 min (Supplementary Table 4) before mounting on Hampton CrystalCap HT loops and cryo-cooling in liquid nitrogen.

Data collection, structure solution and refinement. Diffraction data were collected at the European Synchrotron Radiation Facility, Grenoble, France, with a Mar 225 CCD detector on the microfocus beamline ID23-2 (wavelength, 0.8726 Å) using a 10- μ m focused beam. The microfocus beam was essential for the location of the best diffracting parts of single crystals, as well as allowing several wedges to be collected from different positions. Images were processed with MOSFLM³¹ and SCALA³². The isoprenaline complex was solved by molecular replacement with PHASER³³ using the β_1 AR structure (PDB code 2VT4) as a model¹³. This structure was then used as a starting model for the structure solution of the carmoterol complex. Finally, the carmoterol complex was used as a starting model for both the dobutamine complexes and for the salbutamol complex. Refinement and rebuilding were carried out with REFMAC5³⁴ and COOT³⁵ respectively. The dob92 dobutamine crystal diffracted to a higher resolution (2.5 Å) than the dob102 crystal (2.6 Å), but the dob102 data set was more complete and less anisotropic than dob92 and gave a lower Wilson B factor (Supplementary Table 5). Dictionary entries for the agonists were created using Jligand and PRODRG³⁶. During refinement with REFMAC5 tight non-crystallographic restraints ($\sigma = 0.05$ Å) were applied to the majority (172) of the residues in the two molecules in the asymmetric unit, with their selection based on improvements in R_{free} values. For the salbutamol complex, where the resolution was lower (3.05 Å), all three standard rotamers were modelled for Ser 211 and Ser 215 side chains, and the final choice was made on the basis of the local stereochemistry and features in the difference maps. Hydrogen bond assignments for the ligands were determined using hbplus³⁷ but allowing a maximum hydrogen-acceptor distance of 2.7 Å and a minimum angle of 89 degrees. Superposition of the different complexes was achieved by determining an initial transformation based on the 12 C-terminal residues of helix 2 (90–101) and then using the lsq_imp option of the program O³⁸ to find the largest number of residues that could be superposed without a significant increase in the root mean squared deviation (r.m.s.d.). Cutoff values of between 0.2–0.5 Å for residues to be included in the superposition were found to produce the largest number of residues while maintaining a small r.m.s.d. (<0.15–0.3 Å), depending on the structures being compared. This was repeated using the uppermost residues of helices 3, 6 and 7 to determine the initial transformation, and all cases converged to give the same solution, with 147 residues superposed and a final r.m.s.d. of 0.28 Å for the superposition of the carmoterol and cyanopindolol structures, and lower r.m.s.d. values for superposing different agonist structures on one another. The convergence to a common solution validates this procedure for determining the optimal transformation. Validation of the final refined models was carried out using Molprobity³⁹. Omit densities for the ligands and the surrounding side chains are shown in Supplementary Fig. 3.

The two dobutamine crystals (dob92 and dob102) differed in the crystallization buffer and pH (Supplementary Table 4) and this resulted in slightly different unit cell parameters (Supplementary Table 5) and packing arrangements. The differences between these two structures (overall r.m.s.d. 0.21 Å for monomer A, 0.21 Å for monomer B) provides a measure of the influence of crystal packing forces on the detailed conformation of the receptors. The observed differences in the ligand-binding pocket for monomer B, where there are no direct lattice contacts, emphasizes the conformational flexibility of this region (Supplementary Fig. 6).

Pharmacological analysis of agonist binding to the thermostabilized β_1 AR mutants in whole cells. Stable CHO-K1 cell lines expressing either the wild-type turkey truncated receptor (β trunc), or the β_36 , or the β_6 -m23 or the β_36 -m23 receptors and a CRE-SPAP reporter were used⁴⁰. See Supplementary Table 1 for a description of the constructs. Cells were grown in Dulbecco's modified Eagle's medium nutrient mix F12 (DMEM/F12) containing 10% fetal calf serum and 2 mM D-glutamine in a 37 °C humidified 5% CO₂: 95% air atmosphere.

To analyse the affinities of agonist binding to β_1 AR mutants ³H-CGP 12177 saturation binding and competition binding experiments were performed on whole cells (Supplementary Table 1). Cell lines were grown to confluence in white-sided tissue culture 96-well view plates. ³H-CGP 12177 whole-cell competition binding was performed as previously described⁴⁰ using ³H-CGP 12177 in the range of 0.82–1.80 nM. The K_D values for ³H-CGP 12177 were 0.32 nM (β trunc), 0.85 nM (β_6 -m23), 0.34 nM (β_36) and 0.88 nM (β_36 -m23). For the competition assays, all data points on each binding curve were performed in triplicate and each 96-well plate also contained six determinations of total and non-specific binding. In all cases, the competing ligand completely inhibited the specific binding of ³H-CGP 12177. A one-site sigmoidal response curve was then fitted to the data using GraphPad Prism 2.01 and the IC₅₀ was then determined as the concentration required to inhibit 50% of the specific binding as described previously⁴⁰.

The ability of the receptors to couple to G proteins and induce an increase in cAMP concentrations was determined by measuring the increase in secreted alkaline phosphatase (SPAP) under the transcriptional control of a cAMP response element (CRE). Cells were grown to confluence in clear plastic tissue culture treated 96-well plates and CRE-SPAP secretion into the media measured between 5 and 6 h after the addition of agonist as described previously (Supplementary Fig. 1 and Supplementary Table 3)⁴⁰.

Binding of agonists to β_1 AR mutants expressed in insect cells for structural studies. Receptors β_36 and β_36 -m23 were expressed using the baculovirus expression system in insect cells (High Five) as described previously³⁰. Cells were disrupted by freeze-thaw and membranes prepared by centrifugation. Saturation binding and competition binding experiments were performed using ³H-dihydroalprenolol as described previously⁴¹. Non-specific binding of radioligand to the receptor was determined by including 100 μ M unlabelled alprenolol. The assay mixtures were incubated for 2 h at 30 °C and then filtered on a 96-well glass-fibre filter plates (Millipore) pre-treated with polyethyleneimine. The filters were washed three times with ice-cold buffer (Tris 20 mM pH 8, NaCl 150 mM), dried, and counted in a Beckmann LS 6000 scintillation counter. The apparent IC₅₀ values were determined by nonlinear regression analysis using a one-site competition model in Prism software and K_i values were determined using the Cheng-Prusoff equation⁴².

- Leslie, A. G. W. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).
- McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
- Baker, J. G., Hall, I. P. & Hill, S. J. Agonist actions of “ β -blockers” provide evidence for two agonist activation sites or conformations of the human β_1 -adrenoceptor. *Mol. Pharmacol.* **63**, 1312–1321 (2003).
- Warne, T., Chirnside, J. & Schertler, G. F. Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. *Biochim. Biophys. Acta* **1610**, 133–140 (2003).
- Cheng, Y.-C. & Prusoff, W. H. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).

H2AX prevents CtIP-mediated DNA end resection and aberrant repair in G1-phase lymphocytes

Beth A. Helmink¹, Anthony T. Tubbs¹, Yair Dorsett¹, Jeffrey J. Bednarski^{1,2}, Laura M. Walker¹, Zhihui Feng³, Girdhar G. Sharma³, Peter J. McKinnon⁴, Junran Zhang³, Craig H. Bassing⁵ & Barry P. Sleckman¹

DNA double-strand breaks (DSBs) are generated by the recombination activating gene (RAG) endonuclease in all developing lymphocytes as they assemble antigen receptor genes¹. DNA cleavage by RAG occurs only at the G1 phase of the cell cycle and generates two hairpin-sealed DNA (coding) ends that require nucleolytic opening before their repair by classical non-homologous end-joining (NHEJ)^{1–3}. Although there are several cellular nucleases that could perform this function, only the Artemis nuclease is able to do so efficiently^{2,3}. Here, *in vivo*, we show that in murine cells the histone protein H2AX prevents nucleases other than Artemis from processing hairpin-sealed coding ends; in the absence of H2AX, CtIP can efficiently promote the hairpin opening and resection of DNA ends generated by RAG cleavage. This CtIP-mediated resection is inhibited by γ -H2AX and by MDC-1 (mediator of DNA damage checkpoint 1), which binds to γ -H2AX in chromatin flanking DNA DSBs. Moreover, the ataxia telangiectasia mutated (ATM) kinase activates antagonistic pathways that modulate this resection. CtIP DNA end resection activity is normally limited to cells at post-replicative stages of the cell cycle, in which it is essential for homology-mediated repair^{4,5}. In G1-phase lymphocytes, DNA ends that are processed by CtIP are not efficiently joined by classical NHEJ and the joints that do form frequently use micro-homologies and show significant chromosomal deletions. Thus, H2AX preserves the structural integrity of broken DNA ends in G1-phase lymphocytes, thereby preventing these DNA ends from accessing repair pathways that promote genomic instability.

V(D)J recombination, the reaction that assembles the second exon of all antigen receptor genes, requires the generation and repair of DNA DSBs and occurs exclusively during the G1 phase of the cell cycle¹. The Rag-1 and Rag-2 proteins, which together form an endonuclease referred to as RAG, initiate V(D)J recombination by introducing DNA DSBs at the border of two gene segments and their associated recombination signal (RS) RAG recognition sequences¹. Cleavage by RAG generates a pair of hairpin-sealed coding ends and a pair of blunt signal ends¹. These distinct pairs of DNA ends are processed and joined by NHEJ to form a coding joint and a signal joint, respectively^{2,3}. Signal ends undergo minimal nucleolytic processing before joining. In contrast, hairpin-sealed coding ends must first be opened by an endonuclease and are also frequently processed by exonucleases before joining^{2,3}. This nucleolytic processing results in the antigen receptor gene sequence diversification that is essential for adaptive immunity.

Efficient opening of the hairpin-sealed coding ends generated by RAG cleavage is dependent on the Artemis nuclease^{2,3}. Although other cellular nucleases have enzymatic activity that could open and resect hairpin-sealed coding ends, this functional redundancy is not evident in Artemis-deficient cells; as a result, Artemis-deficient mice and humans are severely lymphopenic³. Thus, the nucleolytic processing

of broken DNA ends in G1-phase lymphocytes must be tightly regulated; however, the cellular components that mediate this regulation are not known. Unrepaired coding ends are resolved as chromosomal translocations at a higher frequency in lymphocytes deficient in both Artemis and H2AX than in lymphocytes with an isolated deficiency of Artemis⁶. Because the formation of these translocations requires the Artemis-independent opening of hairpin-sealed DNA ends, we considered that H2AX might function to restrict the ability of nucleases to act on broken DNA ends in G1-phase lymphocytes.

To test this notion, we generated pre-B-cell lines transformed with the v-abl kinase (hereafter referred to as abl pre-B cells), each of which contained a single integrant of the pMX-DEL^{CJ} retroviral recombination substrate and was wild type (WT:DEL^{CJ}), deficient in Artemis (Artemis^{-/-}:DEL^{CJ}) or deficient in both Artemis and H2AX (Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ})^{6–8}. pMX-DEL^{CJ} had a single pair of recombination signals and rearranged by deletion, resulting in a coding joint that remained within the chromosomal context (Fig. 1a)⁷. Treatment of abl pre-B cells with the v-abl kinase inhibitor STI571 leads to cell cycle arrest in G1 and the induction of RAG, which in NHEJ-deficient cells results in the accumulation of unrepaired coding ends^{6–9}. In G1-arrested Artemis^{-/-}:DEL^{CJ} abl pre-B cells, these unrepaired coding ends were homogeneous in size, as expected given their hairpin-sealed structure (Fig. 1b and Supplementary Figs 1 and 2). In contrast, coding ends in G1-arrested Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} abl pre-B cells were heterogeneous in size and significantly smaller (up to 1 kilobase), suggesting that these hairpin-sealed coding ends had been opened and resected in an Artemis-independent fashion (Fig. 1b and Supplementary Figs 1 and 2). H2AX also regulates coding-end resection in DNA ligase IV-deficient (LigIV^{-/-}) abl pre-B cells, in which hairpin-sealed coding ends can be efficiently opened by Artemis but cannot be ligated (Supplementary Fig. 3). Moreover, blunt chromosomal signal ends generated by RAG cleavage at the pMX-DEL^{SJ} retroviral substrate also show significant resection in LigIV^{-/-}:H2AX^{-/-}:DEL^{SJ} abl pre-B cells (Supplementary Fig. 4)⁷.

To show that hairpin-sealed coding ends have been opened, DNA end structure was assayed by TdT-assisted PCR, which can detect hairpin-opened but not hairpin-sealed coding ends (Supplementary Fig. 5a)¹⁰. In this regard, analysis of Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} abl pre-B cells revealed robust TdT-assisted PCR products, indicating that many of the hairpin-sealed coding ends had been opened (Supplementary Fig. 5c). To quantify the fraction of open coding ends in Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} abl pre-B cells, we performed Southern blot analyses of native and denatured genomic DNA. Denaturing hairpin-opened coding ends dissociates the complementary DNA strands into single-stranded fragments that migrate differently from hairpin-sealed coding ends, whose complementary strands are covalently linked and therefore do not dissociate (Supplementary Fig. 6). Analyses of denatured coding ends from Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ}

¹Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ²Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri 63110, USA. ³Department of Radiation Oncology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁴Department of Genetics and Tumor Cell Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁵Division of Cancer Pathobiology, Department of Pathology and Laboratory Medicine, Center for Childhood Cancer Research, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Abramson Family Cancer Research Institute, Philadelphia, Pennsylvania 19104, USA.

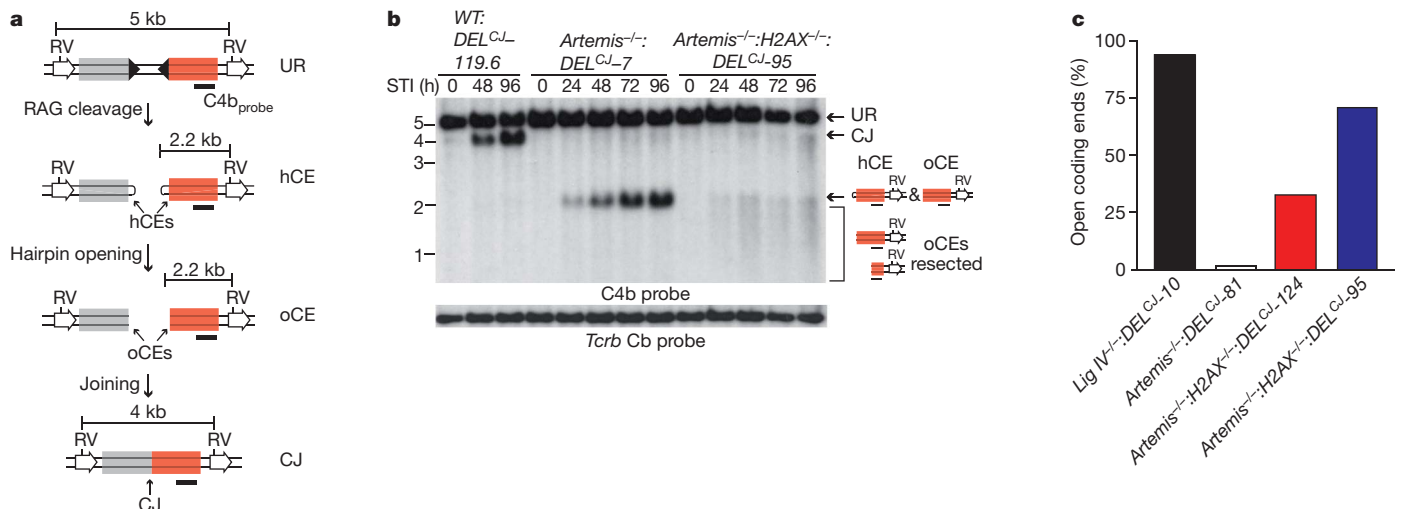


Figure 1 | H2AX inhibits DNA end resection in G1-phase lymphocytes. **a**, pMX-DEL^{CJ} retroviral recombination substrate unarranged (UR), with coding ends that are hairpin-sealed (hCE) or open (oCE), and coding joint (CJ). Recombination signals (filled triangles), EcoRV sites (RV), the C4b probe (black bar) and fragment sizes are shown. **b**, Southern blot analysis of EcoRV-digested genomic DNA from wild-type (WT:DEL^{CJ}-119.6), Artemis^{-/-}:DEL^{CJ} (Artemis^{-/-}:DEL^{CJ}-7) and Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} (Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ}-95) abl pre-B cell lines treated with STI571

(STI). C4b hybridizing UR, CJ and hCEs are indicated, as are oCEs that have been resected (bracket). Tcrb constant region (Cb) probe hybridization was used as a loading control. **c**, Quantification of denaturing Southern blot analyses for open coding ends in LigIV^{-/-}:DEL^{CJ} (LigIV^{-/-}:DEL^{CJ}-10), Artemis^{-/-}:DEL^{CJ} (Artemis^{-/-}:DEL^{CJ}-81) and Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} (Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ}-95 and Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ}-124) abl pre-B cells (see Supplementary Figs 2 and 3 for primary data).

abl pre-B cells revealed that up to 70% of the hairpin-sealed coding ends had been opened (Fig. 1c and Supplementary Fig. 6).

The requirement for H2AX in regulating coding-end processing was also observed at the immunoglobulin light chain (*Igl*)*k* locus in primary G1-phase bone-marrow-derived pre-B cell cultures from Artemis^{-/-} and Artemis^{-/-}:H2AX^{-/-} mice expressing an immunoglobulin heavy chain (*Igh*) transgene (*Igh*t) (Supplementary Fig. 7)⁸. Southern blotting and TdT-assisted PCR revealed that the Jk coding ends in Artemis^{-/-}:H2AX^{-/-}:*Igh*t primary pre-B cells had open hairpins and were resected, whereas those from Artemis^{-/-}:*Igh*t pre-B cells remained hairpin-sealed (Supplementary Fig. 8). Taken together, these data show that H2AX restricts the activity of nucleolytic pathways that would otherwise aberrantly resect unrepaired hairpin-sealed or open coding ends in lymphocytes at the G1-phase of the cell cycle.

H2AX-dependent DNA damage responses generally depend on the phosphorylation of serine 139 by ATM or DNAPKcs to form

γ-H2AX in chromatin flanking DNA DSBs, including those generated by RAG^{11,12}. Reconstitution of LigIV^{-/-}:H2AX^{-/-}:DEL^{CJ} and Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} abl pre-B cells with wild-type H2AX, but not with a serine 139-to-alanine mutant (H2AX^{S139A}), inhibited coding-end resection in these cells (Fig. 2a and Supplementary Fig. 9). These findings implicate γ-H2AX formation at broken DNA ends in maintaining the structure of these ends in G1-phase lymphocytes but do not exclude the possibility that H2AX also inhibits resection through additional pathways that are independent of γ-H2AX formation. Although ATM is required for the optimal formation of γ-H2AX in chromatin flanking RAG DSBs, treatment of Artemis^{-/-}:DEL^{CJ} abl pre-B cells with the ATM kinase inhibitor KU-55933 did not lead to the opening and resection of hairpin-sealed coding ends (Fig. 2b)¹². Rather, treatment of Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} abl pre-B cells with KU-55933 resulted in a significant block in end resection, with a large fraction of DNA ends in these cells being hairpin-sealed (Fig. 2b and

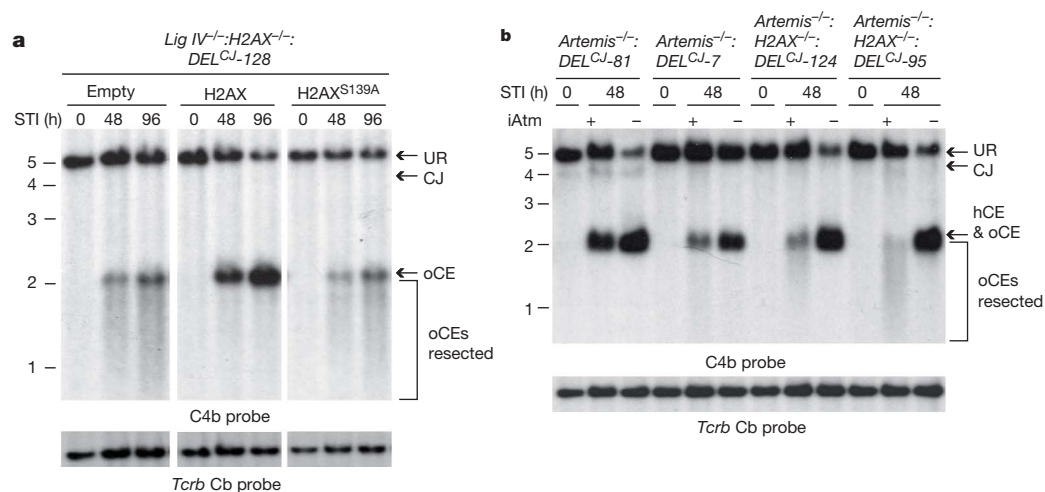


Figure 2 | ATM and γ-H2AX regulate DNA end resection. **a**, pMX-DEL^{CJ} Southern blot analysis (as described in Fig. 1a, b) of STI571-treated LigIV^{-/-}:H2AX^{-/-}:DEL^{CJ} abl pre-B cells reconstituted with an empty retroviral vector (empty) or vectors encoding wild-type H2AX (H2AX) or

H2AX^{S139A}. **b**, pMX-DEL^{CJ} Southern blot analysis of Artemis^{-/-}:DEL^{CJ} and Artemis^{-/-}:H2AX^{-/-}:DEL^{CJ} abl pre-B-cell lines treated with either the ATM inhibitor (iATM) KU55933 (+) or a dimethylsulphoxide vehicle control (-).

data not shown). We conclude that ATM functions to inhibit coding-end resection through the formation of γ -H2AX; however, ATM activity is also required to promote this nucleolytic resection. Thus, ATM modulates the function of antagonistic pathways that both positively and negatively regulate DNA end resection in G1-phase lymphocytes.

We show that, in the absence of H2AX, nucleases other than Artemis can efficiently open and resect hairpin-sealed coding ends in a manner that is dependent on ATM. In this regard, ATM positively regulates CtIP activity in promoting DNA end resection⁵. CtIP binds directly to Nbs1, a component of the Mre11–Rad50–Nbs1 (MRN) complex that associates with RAG DSBs and functions in their repair^{5,11,13–15}. Furthermore, Sae2, the *S. cerevisiae* orthologue of CtIP, functions with Mre11 to promote the opening and resection of hairpin-sealed DNA ends *in vitro*¹⁶. Taken together, these data suggest that H2AX could regulate the ability of CtIP to mediate hairpin opening and DNA end resection in G1-phase lymphocytes. Indeed, knockdown of CtIP in both *Artemis*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} and *Lig IV*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} abl pre-B cells largely blocked the aberrant coding-end resection observed in these cells (Fig. 3a, b and Supplementary Figs 10 and 11). Moreover, in CtIP-deficient *Artemis*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} abl pre-B cells most unrepaired coding ends were hairpin-sealed (Fig. 3c). Although our data show that the hairpin coding ends in *Artemis*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} abl pre-B cells have been opened, we cannot determine the position of opening or the extent of resection after opening. However, it is notable that Sae2 can mediate hairpin opening at significant distances from the hairpin tip *in vitro*¹⁶.

The DNA-damage-response protein MDC-1 is recruited by γ -H2AX to chromatin flanking DNA DSBs^{17,18}. We found that, like γ -H2AX, MDC-1 was also required for the inhibition of the ATM-dependent resection of coding ends in G1-phase lymphocytes (Supplementary Fig. 12). Because CtIP binds to the FHA domain of Nbs1 that also binds MDC-1, this raises the possibility that, on recruitment to DSBs by γ -H2AX, MDC-1 may disrupt CtIP–Nbs1 interactions^{19–23}. The DNA-damage-response protein 53BP1, which is also retained at DSB sites in a γ -H2AX-dependent manner, regulates DNA end resection during V(D)J recombination in thymocytes and during immunoglobulin class switch recombination^{10,24}. In addition, 53BP1 inhibits CtIP-dependent DNA end resection in Brca-1-deficient cells at post-replicative stages of the cell cycle; thus, γ -H2AX may inhibit DNA end resection in these cells by recruiting 53BP1 (ref. 25). However, at post-replicative stages of the cell cycle, H2AX also functions to promote DNA DSB repair by homologous recombination, which requires CtIP-mediated DNA end resection^{5,26}. Thus, H2AX

may function in a cell-cycle-specific manner to modulate the activity of several pathways that regulate DNA end resection.

What is the fate of broken DNA ends processed by CtIP in G1-phase lymphocytes? During V(D)J recombination, Artemis functions primarily to open hairpin-sealed coding ends, after which core NHEJ factors join these DNA ends³. However, hairpin-sealed coding ends opened in a CtIP-dependent manner persisted unrepaired at high levels in *Artemis*^{-/-}:*H2AX*^{-/-} abl pre-B cells (Fig. 1b and Supplementary Fig. 2). In this regard, the single-strand overhangs generated by CtIP-mediated resection during homologous recombination would probably be poor substrates for NHEJ in G1-phase cells⁴. However, this resection could expose regions of homology flanking the DSB, which, if used to mediate DSB repair by homology-driven repair pathways, would form joints with chromosomal deletions⁴. Indeed, PCR and sequence analyses of coding joints formed in *Artemis*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} abl pre-B cells revealed that they were heterogeneous in size, in contrast with those formed in WT:*DEL*^{CJ} or *Artemis*^{-/-}:*DEL*^{CJ} abl pre-B cells (Fig. 4a, b and Supplementary Figs 13 and 14). These joints had significant deletions extending up to 700 base pairs, which is the maximum size deletion that would be amplified by the PCR approach used (Fig. 4a, b and Supplementary Fig. 14). Moreover, the coding joints formed in *Artemis*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} abl pre-B cells used microhomologies at a higher frequency than those formed in WT:*DEL*^{CJ} abl pre-B cells (50% versus 5%; Fig. 4c and Supplementary Fig. 14). Analysis of T-cell receptor β (*Tcrb*) Db1 to Jb1.1/Jb1.2 joints in *Artemis*^{-/-}:*H2AX*^{-/-} thymocytes revealed that these joints similarly showed significant deletions compared with those formed in either *Artemis*^{-/-} or wild-type thymocytes (Supplementary Fig. 15). We conclude that RAG-mediated DNA breaks generated in H2AX-deficient lymphocytes that are processed in a CtIP-dependent manner are resistant to repair by classical NHEJ. However, these DNA ends can be channelled into homology-driven repair pathways, resulting in joints that form significant chromosomal deletions.

The requirement for H2AX in the prevention of CtIP-dependent resection and the resolution of RAG DSBs as chromosomal deletions is congruent with the phenotype of H2AX-deficient mice, which are predisposed to lymphoid tumours that can harbour chromosomal lesions indicative of aberrantly resolved RAG DSBs^{27,28}. However, chromosomal V(D)J recombination proceeds efficiently in H2AX-deficient abl pre-B cells^{6,29}. Moreover, DbJb joints formed in *H2AX*^{-/-} thymocytes did not exhibit significant deletions or an increase in micro-homology usage in comparison with wild-type thymocytes (Supplementary Figs 16 and 17). Thus, H2AX may be required for

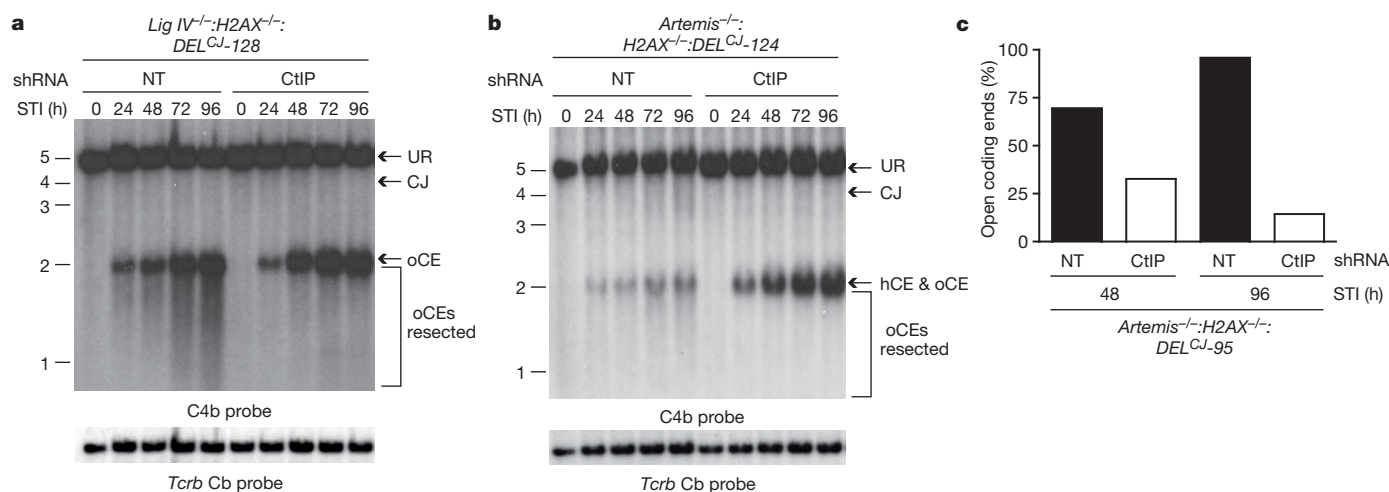


Figure 3 | H2AX prevents CtIP-mediated DNA end resection. **a, b**, pMX-*DEL*^{CJ} Southern blot analysis (as described in Fig. 1a, b) of STI571-treated *Lig IV*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} (**a**) and *Artemis*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} (**b**) abl pre-B cells expressing either non-targeting (NT) or CtIP-specific (CtIP) shRNAs.

(c) Quantification of denaturing Southern blot analysis for open CEs from STI571-treated *Artemis*^{-/-}:*H2AX*^{-/-}:*DEL*^{CJ} abl pre-B cells expressing either NT or CtIP-specific shRNAs.

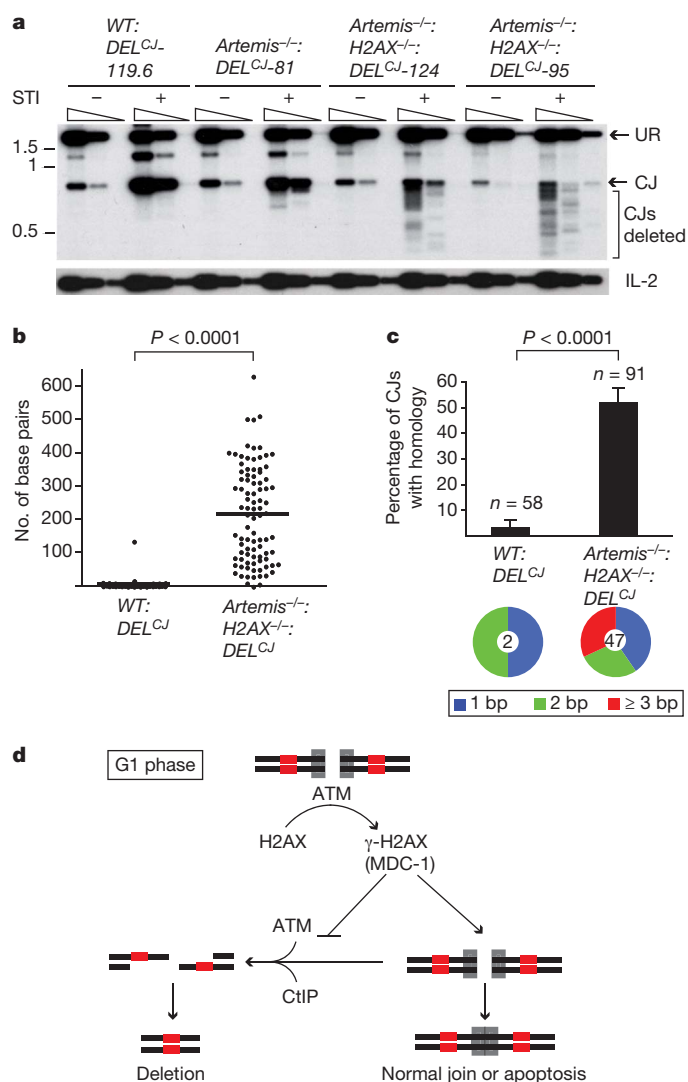


Figure 4 | Aberrant joining in H2AX-deficient cells. **a**, PCR products representing normal and deleted (bracketed) pMX- DEL^{CJ} CJs from WT: DEL^{CJ} , $Artemis^{-/-}$: DEL^{CJ} and $Artemis^{-/-}$: $H2AX^{-/-}$: DEL^{CJ} abl pre-B cells treated with STI571. Serial fivefold dilutions of genomic DNA were amplified. IL-2 gene PCR was used as a loading control. **b**, **c**, Base-pairs deleted (**b**) and microhomology utilization (**c**) in pMX- DEL^{CJ} CJs sequenced from WT: DEL^{CJ} and $Artemis^{-/-}$: $H2AX^{-/-}$: DEL^{CJ} abl pre-B cells (Supplementary Fig. 14). The total number of sequenced coding joints (n) is indicated in **c**. The pie chart shows the fraction of joints with 1, 2 or at least 3 microhomologies and the total number (centre) of joints with microhomologies. **d**, Model for H2AX function in DNA end processing in G1-phase lymphocytes. Red blocks represent homologous sequences. Results are shown as means and s.e.m.; P values were determined by using Student's t -test with Welch's correction for unequal variances.

the repair of a limited subset of RAG DSBs. Alternatively, other proteins may compensate for a more general requirement for H2AX in DNA DSB repair during V(D)J recombination. In agreement with this notion, whereas isolated deficiencies in H2AX or XRCC4-like factor (XLF) have a minimal effect on chromosomal V(D)J recombination, the combined deficiency of H2AX and XLF in abl pre-B cells results in a significant accumulation of unrepaired coding ends²⁹. Moreover, these coding ends are more extensively resected than those in either $Artemis^{-/-}$: $H2AX^{-/-}$ or $Lig IV^{-/-}$: $H2AX^{-/-}$ abl pre-B cells, raising the possibility that H2AX and XLF have overlapping activities in modulating DNA end resection²⁹.

We have shown that H2AX maintains the structural integrity of broken DNA ends generated during V(D)J recombination in G1-phase lymphocytes. This protective function of H2AX would ensure

that unrepaired DNA ends are either joined through classical NHEJ or signal for elimination of the cell by apoptosis (Fig. 4d). In the absence of H2AX, ATM/CtIP-dependent resection creates genomic instability by allowing DNA ends to be shuttled into homology-driven repair pathways that can form potentially dangerous chromosomal deletions (Fig. 4d). Formation of chromosomal translocations in H2AX-deficient mice may rely on these defects in DNA end processing coupled with the diminished retention of coding ends in post-cleavage complexes in H2AX-deficient cells⁶. Significant parallels can be drawn between mechanisms that protect, process and repair RAG DSBs and those incurred by environmental genotoxins. In this regard, a subset of genotoxic DSBs require Artemis for their repair by NHEJ, suggesting that these broken DNA ends must undergo nucleolytic processing³⁰. Furthermore, the repair of these genotoxic DSBs may also depend on H2AX³⁰. Thus, our finding that H2AX regulates the processing of unrepaired DNA ends during V(D)J recombination may reflect a broader function of H2AX in regulating the nucleolytic processing of DNA DSBs generated during other physiological processes and by genotoxic agents.

METHODS SUMMARY

Abl pre-B cell lines containing the pMX- DEL^{CJ} retroviral recombination substrate were generated and maintained, and interleukin (IL)-7-dependent pre-B cells were cultured as described previously^{7,8}. Standard protocols for Southern blotting (native and denaturing), western blotting, flow cytometry, retroviral-mediated protein expression and short hairpin RNA (shRNA)-mediated knockdown were followed. TdT-assisted PCR for coding ends generated during rearrangement of pMX- DEL^{CJ} and the *Igk* locus was performed as described in Methods¹⁰.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 June; accepted 15 October 2010.

Published online 15 December 2010.

1. Fugmann, S. D., Lee, A. I., Shockett, P. E., Villy, I. J. & Schatz, D. G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* **18**, 495–527 (2000).
2. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination. *DNA Repair (Amst.)* **3**, 817–826 (2004).
3. Rooney, S., Chaudhuri, J. & Alt, F. W. The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol. Rev.* **200**, 115–131 (2004).
4. Huertas, P. DNA resection in eukaryotes: deciding how to fix the break. *Nature Struct. Mol. Biol.* **17**, 11–16 (2010).
5. You, Z. & Bailis, J. M. DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. *Trends Cell Biol.* **20**, 402–409 (2010).
6. Yin, B. *et al.* Histone H2AX stabilizes broken DNA strands to suppress chromosome breaks and translocations during V(D)J recombination. *J. Exp. Med.* **206**, 2625–2639 (2009).
7. Bredemeyer, A. L. *et al.* ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466–470 (2006).
8. Bredemeyer, A. L. *et al.* DNA double-strand breaks activate a multi-functional genetic program in developing lymphocytes. *Nature* **456**, 819–823 (2008).
9. Muljo, S. A. & Schlissel, M. S. A small molecule Abl kinase inhibitor induces differentiation of Abelson virus-transformed pre-B cell lines. *Nature Immunol.* **4**, 31–37 (2003).
10. Difilippantonio, S. *et al.* 53BP1 facilitates long-range DNA end-joining during V(D)J recombination. *Nature* **456**, 529–533 (2008).
11. Chen, H. T. *et al.* Response to RAG-mediated VDJ cleavage by NBS1 and γ -H2AX. *Science* **290**, 1962–1965 (2000).
12. Savic, V. *et al.* Formation of dynamic γ -H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin. *Mol. Cell* **34**, 298–310 (2009).
13. Chen, L., Nievera, C. J., Lee, A. Y. & Wu, X. Cell cycle-dependent complex formation of BRCA1/CtIP/MRN is important for DNA double-strand break repair. *J. Biol. Chem.* **283**, 7713–7720 (2008).
14. Deriano, L., Stracker, T. H., Baker, A., Petrini, J. H. & Roth, D. B. Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. *Mol. Cell* **34**, 13–25 (2009).
15. Helmink, B. A. *et al.* MRN complex function in the repair of chromosomal RAG-mediated DNA double-strand breaks. *J. Exp. Med.* **206**, 669–679 (2009).
16. Lengsfeld, B. M., Rattray, A. J., Bhaskara, V., Ghirlando, R. & Paull, T. T. Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. *Mol. Cell* **28**, 638–651 (2007).
17. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213–1226 (2005).

18. Lou, Z. *et al.* MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol. Cell* **21**, 187–200 (2006).
19. Melander, F. *et al.* Phosphorylation of SPT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage-modified chromatin. *J. Cell Biol.* **181**, 213–226 (2008).
20. Chapman, J. R. & Jackson, S. P. Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. *EMBO Rep.* **9**, 795–801 (2008).
21. Spycher, C. *et al.* Constitutive phosphorylation of MDC1 physically links the MRE11–RAD50–NBS1 complex to damaged chromatin. *J. Cell Biol.* **181**, 227–240 (2008).
22. Lloyd, J. *et al.* A supramodular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA damage. *Cell* **139**, 100–111 (2009).
23. Williams, R. S. *et al.* Nbs1 flexibly tethers Ctp1 and Mre11–Rad50 to coordinate DNA double-strand break processing and repair. *Cell* **139**, 87–99 (2009).
24. Bothmer, A. *et al.* 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J. Exp. Med.* **207**, 855–865 (2010).
25. Bunting, S. F. *et al.* 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243–254 (2010).
26. Xie, A. *et al.* Distinct roles of chromatin-associated proteins MDC1 and 53BP1 in mammalian double-strand break repair. *Mol. Cell* **28**, 1045–1057 (2007).
27. Bassing, C. H. *et al.* Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. *Cell* **114**, 359–370 (2003).
28. Celeste, A. *et al.* H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. *Cell* **114**, 371–383 (2003).
29. Zha, S. *et al.* XLF has redundant functions with ATM and H2AX in V(D)J recombination and non-homologous DNA end-joining. *Nature* doi:10.1038/nature09604 (this issue).
30. Riballo, E. *et al.* A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to γ -H2AX foci. *Mol. Cell* **16**, 715–724 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Oltz and F. Alt for critical review of the manuscript; R. Baer for providing us with the anti-CtIP antibodies; and G. Longmore and Y. Feng for pFLRU, pHR Δ 8.2R and pCMV-VSVg. This work is supported by National Institutes of Health (NIH) grants AI074953 (B.P.S.), AI47829 (B.P.S.), CA136470 (B.P.S. and C.H.B.), CA125195 (C.H.B.), CA21765 (P.J.McK.) and NS37956 (P.J.McK.). J.J.B. is supported by a NIH Ruth L. Kirschstein National Research Service Award (T32 HD007499) and a Children's Discovery Institute Fellows Award. C.H.B. was a Pew Scholar in the Biomedical Sciences program and is a Leukemia and Lymphoma Society Scholar.

Author Contributions B.P.S. and B.A.H. conceived the study and wrote the paper. B.P.S. and C.H.B. designed experiments and interpreted data. B.A.H., A.T.T., Y.D., G.S., J.Z. and J.J.B. designed and performed experiments and interpreted data. Z.F. and L.M.W. performed experiments. P.J.McK. provided important reagents.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.P.S. (sleckman@wustl.edu).

METHODS

Mice. All animals were housed in a specific pathogen-free facility at Washington University School of Medicine, and all animal protocols were approved by the Washington University Institutional Animal Care and Use Committee.

Cell line generation and cell culture. *Rag*^{-/-}, *Artemis*^{-/-} and WT abl pre-B cells were described previously^{7,8}. *Artemis*^{-/-}:*H2AX*^{-/-} v-abl-transformed pre-B cells were generated as described previously⁶. *LigIV*^{-/-} and *LigIV*^{-/-}:*H2AX*^{-/-} abl pre-B cells were generated by treating *LigIV*^{loxp/loxp} and *LigIV*^{loxp/loxp}:*H2AX*^{-/-} abl pre-B cells with a Tat-Cre fusion protein^{6,31}. Cells were incubated for 1 h in medium containing 50 µg ml⁻¹ Tat-Cre and subcloned 48 h later. Cre-mediated deletion of the *DNA Ligase IV* gene was confirmed by PCR and Southern blotting. All lines were transduced with pMX-DEL^{CJ} or pMX-DEL^{SJ} by co-centrifugation, and clonal populations each containing a single integrant of the retroviral substrate were isolated by limiting dilution as described previously⁸. Cells were treated with 3 µM STI571 (Novartis) for the indicated durations at a concentration of 10⁶ cells ml⁻¹. KU-55933 (Tocris) was used at a concentration of 15 µM. Primary bone-marrow pre-B-cell cultures were generated by harvesting bone marrow from *Rag*^{-/-}:*Ightg*, *Artemis*^{-/-}:*Ightg* or *Artemis*^{-/-}:*H2AX*^{-/-}:*Ightg* mice followed by culture for 6–10 days in medium containing 5 ng ml⁻¹ IL-7 maintained at a concentration of about 2 × 10⁶ cells ml⁻¹ before withdrawal of IL-7 (ref. 8).

Retroviral reconstitution and lentiviral knockdown. Reconstitution of *Artemis*^{-/-}:*H2AX*^{-/-} and *LigIV*^{-/-}:*H2AX*^{-/-} abl pre-B cells was performed by retroviral transduction with either empty retrovirus or retrovirus containing cDNAs encoding H2AX or H2AX^{S139A}. A cDNA encoding H2AX^{S139A} (serine 139 changed to alanine) was generated by PCR-based site-directed mutagenesis of WT H2AX cDNA. cDNAs encoding WT H2AX and H2AX^{S139A} were cloned into the pMX-PIE retroviral vector and cells were transduced by co-centrifugation as described previously⁸. Cells expressing the retroviral construct were obtained by flow cytometric cell sorting of cells expressing GFP using a FACSVantage (BD Biosciences). Generation of lentiviral shRNAs vectors was performed with the previously described pFLRU:YFP lentiviral vector^{32,33}. CtlP-specific and non-targeting (NT) shRNAs were cloned into the pFLRU:YFP lentiviral vector. Sequences targeted by the shRNA were CtlP (5'-GAGCAGACCTTTCTCAGTA-3') and NT (5'-GGTTCGATGTCCCAATTCTG-3'). These pFLRU:shRNA:YFP vectors (4 µg) were individually co-transfected with 4 µg of pHR^Δ8.2R packaging vector and 1 µg of pCMV-VSVg envelope plasmid into HEK293T cells plated at 70% confluence in 6-cm² plates, using Lipofectamine 2000 (Invitrogen). Medium was replaced 12 h after transfection. Supernatants were harvested 24 h later. Transduction of abl pre-B cells was performed by co-centrifugation with viral supernatant at 1,800 r.p.m. (650g) for 90 min, with Polybrene added to 5 µg ml⁻¹. Cells expressing the pFLRU-shRNA vectors were obtained by flow cytometric cell sorting of cells expressing YFP, with a FACSVantage (BD Biosciences).

Southern blotting analyses. Standard Southern blot analysis of pMX-DEL^{CJ} and pMX-DEL^{SJ} was performed with *EcoRV*-digested genomic DNA and the C4b probe as described previously⁸. Southern blot analyses of coding ends generated during rearrangement at the *Igk* locus were performed as described previously on genomic DNA digested with *SacI* and *EcoRI*, with the JkIII probe⁷. Denaturing agarose-gel electrophoresis was performed as described previously, with modifications³⁴. In brief, 40 µg of genomic DNA was digested overnight with *EcoRV* in a 400-µl volume and concentrated to 30 µl. The DNA was then resuspended with the addition of 5 volumes of a solution containing 8 M urea, 1% Nonidet P40, 1 mM Tris-HCl pH 8.0 and 0.5 mg ml⁻¹ bromophenol blue. This DNA solution was divided in two; one half was heated at 90 °C for 8 min to denature the genomic DNA, and the other half was incubated on ice. After 8 min, the heated DNA samples were placed on ice before electrophoresis at 50 V and 4 °C on a 1.2% agarose Tris-acetate-EDTA gel with 1 M urea for about 24 h in TAE buffer also

containing 1 M urea. Quantification was performed with ImageJ software. To calculate the percentage of hairpin-sealed coding ends, we measured the integrated density (ID) of the band representing open coding ends (oCEs) and that representing hairpin-sealed coding ends (hCEs) and subtracted background ID levels to obtain the corrected ID levels for each (corr. ID^{oCE} and corr. ID^{hCE}, respectively). The ID for the closed ends was divided by the sum of the IDs of closed and open ends and converted to a percentage: hCE (%) = 100(corr. ID^{hCE})/(corr. ID^{hCE} + corr. ID^{oCE}). **TdT-assisted PCR.** TdT-assisted PCR analysis of pMX-DEL^{CJ} coding ends was performed as described previously, with the IRES REV5 oligonucleotide (5'-CTCGACTAAACACATGTAAAGC-3') for the primary PCR reaction, the IRES REV4 oligonucleotide (5'-CCCTTGTGTGAATACGCTTG-3') for the secondary PCR reaction and the I4 oligonucleotide (5'-TAAGATACACCTGCAAAGGCG-3') as a probe¹⁰. Similar conditions were used for PCR analyses of coding ends generated during rearrangement of the endogenous *Igk* locus, using the Jk2 ds oligonucleotide (5'-CCACAAGAGGTTGGAATGATTTC-3') for amplification and the Jk oligonucleotide (5'-GTAGTCTTCTCAACTCTGTTCAC-3') as a probe. *IL-2* gene PCR, which was provided as a DNA loading control, was performed as described previously¹⁰.

Analysis of pMX-DEL^{CJ} coding joints. pMX-DEL^{CJ} coding joints were amplified by using oligonucleotides pC (5'-GCACGAAGTCTTGAGACCT-3') and IRES REV5 (5'-CTCGACTAAACACATGTAAAGC-3'). Genomic DNA (300 ng) was used in the original amplification with serial fivefold dilutions. Oligonucleotide IR4 (5'-CCCTTGTGTGAATACGCTTG-3') was used as a probe. Cloning and sequencing of pMX-DEL^{CJ} coding joints was performed as described previously¹⁵. *P* values for both Fig. 4b and Fig. 4c were calculated by Student's *t*-test with Welch's correction for unequal variances.

Analysis of *Tcrb* coding joints. Db1-Jb1.1 and Db1-Jb1.2 CJs were amplified by using oligonucleotides Db1 us (5'-CCTTCCTTATCTTCAACTC-3') and Jb1.2 ds (5'-CCTGACTTCCACCCGAGGTT-3') with the following PCR parameters: 92 °C for 1 min 30 s, 55 °C for 2 min 30 s, 72 °C for 1 min. Thymic DNA (300 ng) was used in the original amplification (13 cycles). PCR products were digested with *Bgl*III for 3 h before a second round of amplification (30 cycles) to permit the isolation of DJb CJs without the competing amplification from the germline *Tcrb* locus. Southern blot analyses of the PCR products were performed with oligonucleotide Jb (5'-GTAATCAGAGGAAGGATG-3') as a probe. Thymic DNA isolated from three individual mice for each genotype (*WT*, *Artemis*^{-/-} and *Artemis*^{-/-}:*H2AX*^{-/-}) was analysed. Db1-Jb1.1 CJs from *WT* and *H2AX*^{-/-} thymic DNA were cloned and sequenced.

Flow cytometric analyses. Flow cytometric analyses were performed on a FACSCaliber (BD Biosciences) with fluorescein isothiocyanate-conjugated anti-CD45R/B220 and allophycocyanin-conjugated IgM and the appropriate isotype control. All antibodies were from BD Biosciences. Cell cycle analyses were performed by assessing DNA content by incubation with Hoechst 33342 dye (Invitrogen) for 1 h at 37 °C before flow cytometric analysis. The percentage of cells in G1 and S + G2 + M were approximated with the Dean-Jett-Fox method in FLOW-JO 8.8.6 software.

- Shull, E. R. *et al.* Differential DNA damage signaling accounts for distinct neural apoptotic responses in ATLD and NBS. *Genes Dev.* **23**, 171–180 (2009).
- Lu, L. *et al.* Actin stress fiber pre-extension in human aortic endothelial cells. *Cell Motil. Cytoskeleton* **65**, 281–294 (2008).
- Langer, E. M. *et al.* Ajuba LIM proteins are snail/slug corepressors required for neural crest development in *Xenopus*. *Dev. Cell* **14**, 424–436 (2008).
- Hegedus, E., Kokai, E., Kotlyar, A., Dombradi, V. & Szabo, G. Separation of 1–23-kb complementary DNA strands by urea-agarose gel electrophoresis. *Nucleic Acids Res.* **37**, e112 (2009).

ATM damage response and XLF repair factor are functionally redundant in joining DNA breaks

Shan Zha^{1*†}, Chunguang Guo^{1*}, Cristian Boboila¹, Valentyn Oksenysh¹, Hwei-Ling Cheng¹, Yu Zhang¹, Duane R. Wesemann¹, Grace Yuen¹, Harin Patel¹, Peter H. Goff¹, Richard L. Dubois² & Frederick W. Alt¹

Classical non-homologous DNA end-joining (NHEJ) is a major mammalian DNA double-strand-break (DSB) repair pathway. Deficiencies for classical NHEJ factors, such as XRCC4, abrogate lymphocyte development, owing to a strict requirement for classical NHEJ to join V(D)J recombination DSB intermediates^{1,2}. The XRCC4-like factor (XLF; also called NHEJ1) is mutated in certain immunodeficient human patients and has been implicated in classical NHEJ^{3–6}; however, XLF-deficient mice have relatively normal lymphocyte development and their lymphocytes support normal V(D)J recombination⁵. The ataxia telangiectasia-mutated protein (ATM) detects DSBs and activates DSB responses by phosphorylating substrates including histone H2AX⁷. However, ATM deficiency causes only modest V(D)J recombination and lymphocyte developmental defects, and H2AX deficiency does not have a measurable impact on these processes^{7–9}. Here we show that XLF, ATM and H2AX all have fundamental roles in processing and joining DNA ends during V(D)J recombination, but that these roles have been masked by unanticipated functional redundancies. Thus, combined deficiency of ATM and XLF nearly blocks mouse lymphocyte development due to an inability to process and join chromosomal V(D)J recombination DSB intermediates. Combined XLF and ATM deficiency also severely impairs classical NHEJ, but not alternative end-joining, during IgH class switch recombination. Redundant ATM and XLF functions in classical NHEJ are mediated by ATM kinase activity and are not required for extra-chromosomal V(D)J recombination, indicating a role for chromatin-associated ATM substrates. Correspondingly, conditional H2AX inactivation in XLF-deficient pro-B lines leads to V(D)J recombination defects associated with marked degradation of unjoined V(D)J ends, revealing that H2AX has a role in this process.

Assembly of immunoglobulin and T-cell-receptor variable region exons is initiated by the RAG1 and RAG2 endonuclease (hereafter referred to as RAG), which generates DNA DSBs between a pair of participating V, D, or J coding segments and flanking recombination signal sequences¹⁰. V(D)J recombination is completed via joining, respectively, of the two coding segments and two recombination signal sequences by classical NHEJ². Whereas XLF-deficient ($XLF^{\Delta/\Delta}$) embryonic stem cells and mouse embryonic fibroblasts are impaired for V(D)J recombination of extra-chromosomal substrates⁵, $XLF^{\Delta/\Delta}$ mice are only modestly impaired for lymphocyte development, and $XLF^{\Delta/\Delta}$ pro-B lines, although having increased sensitivity to ionizing radiation (IR), perform nearly normal V(D)J recombination⁵. Thus, unknown factors may compensate for XLF V(D)J recombination functions in developing lymphocytes⁵. Among the candidates, we considered ATM, which is activated by RAG-generated DSBs^{7,8,11}. To elucidate whether ATM has an overlapping V(D)J joining function with XLF, we bred $XLF^{\Delta/\Delta}$ mice⁵ with ATM-deficient ($Atm^{-/-}$)¹² mice to generate $XLF^{\Delta/\Delta}Atm^{-/-}$ mice. $XLF^{\Delta/\Delta}Atm^{-/-}$ mice were live born but

were significantly smaller than control littermates (Supplementary Fig. 1).

$XLF^{\Delta/\Delta}$ and $Atm^{-/-}$ mice had only a modest (2–3-fold) reduction in thymocyte numbers and no gross alterations in thymocyte development, as revealed by staining for the CD4 and CD8 differentiation markers (Fig. 1a, b and Supplementary Table 1). In contrast, $XLF^{\Delta/\Delta}Atm^{-/-}$ mice had a greater than 20-fold decrease in thymocyte numbers, to levels nearly as low as those of $Rag2^{-/-}$ mice, with an overall developmental pattern reminiscent of that of certain classical NHEJ deficient mice with a 'leaky' V(D)J recombination block². B-cell development was also relatively unimpaired in XLF- and ATM-deficient mice, with both having only modestly reduced (2–3-fold) B220⁺IgM⁺ splenic B-cell numbers (Fig. 1a, c and Supplementary Table 1)^{5,12}. In contrast, $XLF^{\Delta/\Delta}Atm^{-/-}$ mice had extremely low splenic B-cell numbers (Fig. 1a, c and Supplementary Table 1). Analyses of bone marrow B-cell development in $XLF^{\Delta/\Delta}Atm^{-/-}$ mice suggested an impairment at the CD43⁺B220⁺ progenitor (pro-) B-cell stage in which V(D)J recombination is initiated, as shown by the near absence of B220⁺CD43⁺ precursor B cells (Fig. 1a). To test further whether impaired B-cell development in $XLF^{\Delta/\Delta}Atm^{-/-}$ mice involved a V(D)J recombination defect, we bred IgH and IgL loci that contained knock-in mutations of pre-assembled IgH and IgL variable region exons (referred to as HL)¹³ into the $XLF^{\Delta/\Delta}Atm^{-/-}$ background and found a significant rescue of B-cell, but not T-cell, development (Fig. 1a–c and Supplementary Table 1). Together, these findings indicate that XLF/ATM double-deficiency severely impairs T- and B-cell development by impairing V(D)J recombination.

To test unequivocally for V(D)J recombination end-joining defects, we generated $v-abl$ transformed pro-B-cell lines from wild-type, $XLF^{\Delta/\Delta}$, $Atm^{-/-}$ and $XLF^{\Delta/\Delta}Atm^{-/-}$ mice that carried *Bcl-2* transgenes⁸. Treatment of $v-abl$ transformed pro-B lines with STI571, a *v-abl* kinase inhibitor, arrests cells in G1 and induces RAG, leading to efficient V(D)J recombination of integrated substrates in wild-type cells⁸. The *Bcl-2* transgene obviates apoptotic effects of STI571 (ref. 8). We generated multiple pro-B lines from each genotype that harboured, respectively, either a V(D)J recombination substrate designed to assay coding joins (CJs) and unjoined coding ends (CEs) (Fig. 2a) or a substrate designed to assay recombination signal sequence joins (SJs) and unjoined recombination signal sequence ends (SEs) (Fig. 2b). For these experiments, DNA from individual lines was prepared at day 0 (before treatment), day 2 and day 4 of STI571 treatment, digested with restriction endonucleases and assayed for hybridization to the indicated probes (Fig. 2c, d). Wild-type and $XLF^{\Delta/\Delta}$ lines generated substantial CJ and SJ levels at day 2 and 4 with little or no obvious free CEs, indicative of a classical NHEJ defect (Fig. 2c, d). $Atm^{-/-}$ lines also generated substantial levels of CJs and SJs, but, consistent with previous studies⁸, also generated a modest level of unjoined CEs at day 2 that appeared partially resolved by day 4 (Fig. 2c). However, there was

¹Howard Hughes Medical Institute, The Children's Hospital, the Immune Disease Institute and the Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Pathology and Cell Biology, Department of Pediatrics, Institute for Cancer Genetics, Columbia University, New York City, New York 10032, USA. [†]Present address: Department of Pathology and Cell Biology, Department of Pediatrics, Institute for Cancer Genetics, Columbia University, New York City, New York 10032, USA.

*These authors contributed equally to this work.

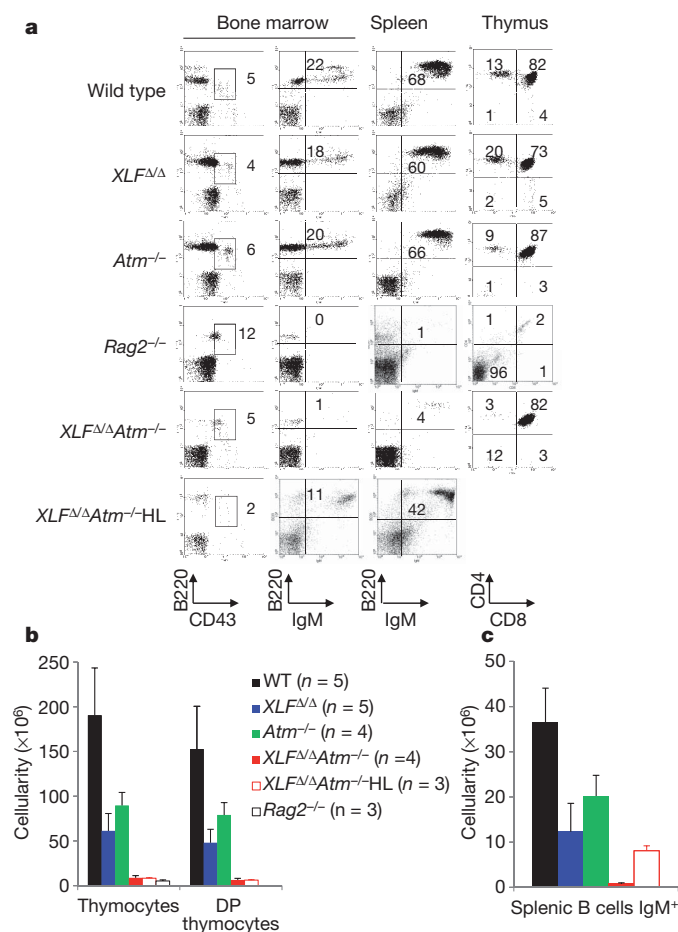


Figure 1 | ATM and XLF have redundant functions in lymphocyte development. **a**, Representative flow cytometric analyses of bone marrow, spleen and thymus from wild-type, $XLF^{\Delta/\Delta}$, $Atm^{-/-}$, $Rag2^{-/-}$, $XLF^{\Delta/\Delta}Atm^{-/-}$ and $XLF^{\Delta/\Delta}Atm^{-/-}HL$ mice (see Methods for further description of mouse lines). Numbers on the plot are percentages of total cells represented by the indicated population. **b**, **c**, Total thymocyte and CD4⁺CD8⁺ double positive (DP) thymocyte numbers (**b**) and IgM⁺ splenic B-cell numbers (**c**). Each value listed represents the average \pm standard deviation from at least three mice between 4–12 weeks of age. See Supplementary Table 1 for details.

no obvious recombination signal sequence joining defect in the $Atm^{-/-}$ lines (Fig. 2d). In contrast, $XLF^{\Delta/\Delta}Atm^{-/-}$ lines had little accumulation of CJs or SJs at either time point and, instead, accumulated unjoined CE and SEs, respectively (Fig. 2c, d).

We also tested for V(D)J recombination defects with a substrate that activates a GFP gene upon inversional V(D)J recombination and which, via Southern blotting, reveals CJs, hybrid joins (aberrant joins in which a recombination signal sequence is fused to a coding end) and free CEs (Fig. 2e, f). We clonally integrated a single-copy inversional V(D)J substrate into $XLF^{\Delta/\Delta}$ pro-B lines that were also homozygous for a conditional knockout ATM allele ($Atm^{C/C}$)¹² and then deleted floxed $Atm^{C/C}$ sequences via Cre recombinase to generate $XLF^{\Delta/\Delta}Atm^{-/-}$ lines with the same substrate integration. Thus, these matched sets of lines allow assay of a given integrated substrate in $XLF^{\Delta/\Delta}$ pro-B lines before and after elimination of ATM. We treated inversional substrate containing wild-type, $Atm^{-/-}$, $XLF^{\Delta/\Delta}Atm^{C/C}$, $XLF^{\Delta/\Delta}Atm^{-/-}$ and $Xrcc4^{-/-}$ lines with STI571 and assayed for V(D)J recombination both by GFP expression (Supplementary Fig. 2) and Southern blotting (Fig. 2f). Both assays confirmed the severe V(D)J recombination defect in $XLF^{\Delta/\Delta}Atm^{-/-}$ pro-B lines and Southern blotting confirmed severely defective end-joining, as revealed by a marked decrease in CJs and a marked increase in unjoined CEs (Fig. 2f). The severity of the inversional V(D)J joining defect in $XLF^{\Delta/\Delta}Atm^{-/-}$ pro-B lines was

similar to that of XRCC4-deficient pro-B lines (Fig. 2f and Supplementary Fig. 2). $XLF^{\Delta/\Delta}$ pro-B lines treated with an ATM kinase inhibitor also showed a severe end-joining defect during V(D)J recombination, indicating that the ATM-mediated V(D)J joining activity revealed in XLF-deficient lines is mediated by ATM kinase activity (Fig. 2f). Finally, STI571-treated $XLF^{\Delta/\Delta}Atm^{-/-}$ pro-B lines accumulated unrepaired V(D)J recombination-associated breaks within their endogenous Igk locus, similar to those observed in $Artemis^{-/-}$ (also called $Dclre1c$) pro-B lines, confirming that the V(D)J recombination defect associated with combined XLF and ATM deficiency extends to this endogenous immunoglobulin locus (Supplementary Fig. 3).

To characterize further the V(D)J recombination defect in $XLF^{\Delta/\Delta}Atm^{-/-}$ versus wild-type, $XLF^{\Delta/\Delta}$, $Atm^{-/-}$ and $Xrcc4^{-/-}$ pro-B lines, we assayed for V(D)J recombination on transiently introduced extra-chromosomal substrates¹⁴. As this assay is semi-quantitative, within perhaps a fivefold range, and is best for revealing profound defects, we performed at least four independent assays for each genotype (Supplementary Table 1). As expected^{2,5}, transient coding and recombination signal sequence joining activity for XRCC4-deficient cells was more than 50-fold less than that of wild-type cells, whereas coding and recombination signal sequence joining activity for $XLF^{\Delta/\Delta}$ and $Atm^{-/-}$ cells approached the wild-type range (Supplementary Table 2). Surprisingly, the range of coding and recombination signal sequence joining activity of $XLF^{\Delta/\Delta}Atm^{-/-}$ pro-B lines, although potentially modestly decreased, overlapped that of wild-type and single mutant cells (Supplementary Table 2). Thus, in contrast to severe defects in chromosomal V(D)J recombination, XLF/ATM double mutant pro-B lines lack severe defects in extra-chromosomal V(D)J recombination.

The chromosomal V(D)J joining defect in $XLF^{\Delta/\Delta}Atm^{-/-}$ pro-B lines can be attributed to impaired classical NHEJ, as V(D)J recombination exclusively uses this pathway¹⁵. However, the question remains as to whether $XLF^{\Delta/\Delta}Atm^{-/-}$ cells have more general classical NHEJ defects and whether they are impaired for other forms of end-joining. IgH class switch recombination (CSR) involves the introduction of DSBs into switch (S) region upstream of the C μ constant region exons and their joining to DSBs within a downstream S region, resulting in IgH CSR¹⁶. Although classical NHEJ is a major CSR joining pathway, CSR occurs at reduced levels in classical NHEJ-deficient cells via alternative end-joining (A-EJ)¹³. To assay CSR, we activated wild-type, $XLF^{\Delta/\Delta}$, $Atm^{-/-}$ and $XLF^{\Delta/\Delta}Atm^{-/-}$ HL B cells for 4 days with anti-CD40 and interleukin (IL)-4 to stimulate CSR to IgG1. As expected^{15,17–20}, $XLF^{\Delta/\Delta}$ and $Atm^{-/-}$ B cells switched to IgG1 at about 40% of wild-type levels^{5,17–20}. Moreover, $XLF^{\Delta/\Delta}Atm^{-/-}$ HL B cells also showed substantial residual IgG1 CSR that was on average about 25% of wild-type levels (Fig. 3a, b and Supplementary Fig. 4). To gain further insight into involved joining pathways, we sequenced the S μ to S γ 1 junctions. Classical NHEJ generates CSR junctions with no microhomology (for example, direct joins) and junctions with short (1–2 bp) microhomologies¹³, whereas A-EJ generates CSR junctions that predominantly contain microhomologies¹³. As expected, about 40% of wild-type junctions were direct¹³, whereas only about 22% and 13%, respectively, of $Atm^{-/-}$ and $XLF^{\Delta/\Delta}$ junctions were direct^{5,20}, indicating some classical NHEJ impairment in these mutant B cells (Fig. 3c and Supplementary Fig. 4). However, only about 5% of $XLF^{\Delta/\Delta}Atm^{-/-}$ CSR joins were direct (Fig. 3c and Supplementary Fig. 4), consistent with most of their residual CSR being carried out by A-EJ. These results indicate that combined XLF and ATM deficiency impairs general classical NHEJ during CSR, but does not substantially impair A-EJ.

Our finding that the overlapping function of ATM with XLF involves ATM kinase activity and is required for chromosomal versus extra-chromosomal V(D)J joining indicates that this function involves ATM substrates. In response to DSBs, ATM phosphorylates H2AX⁷. However, H2AX deficiency is not known to have a detectable impact on V(D)J recombination^{9,21}. To test for overlapping H2AX and XLF

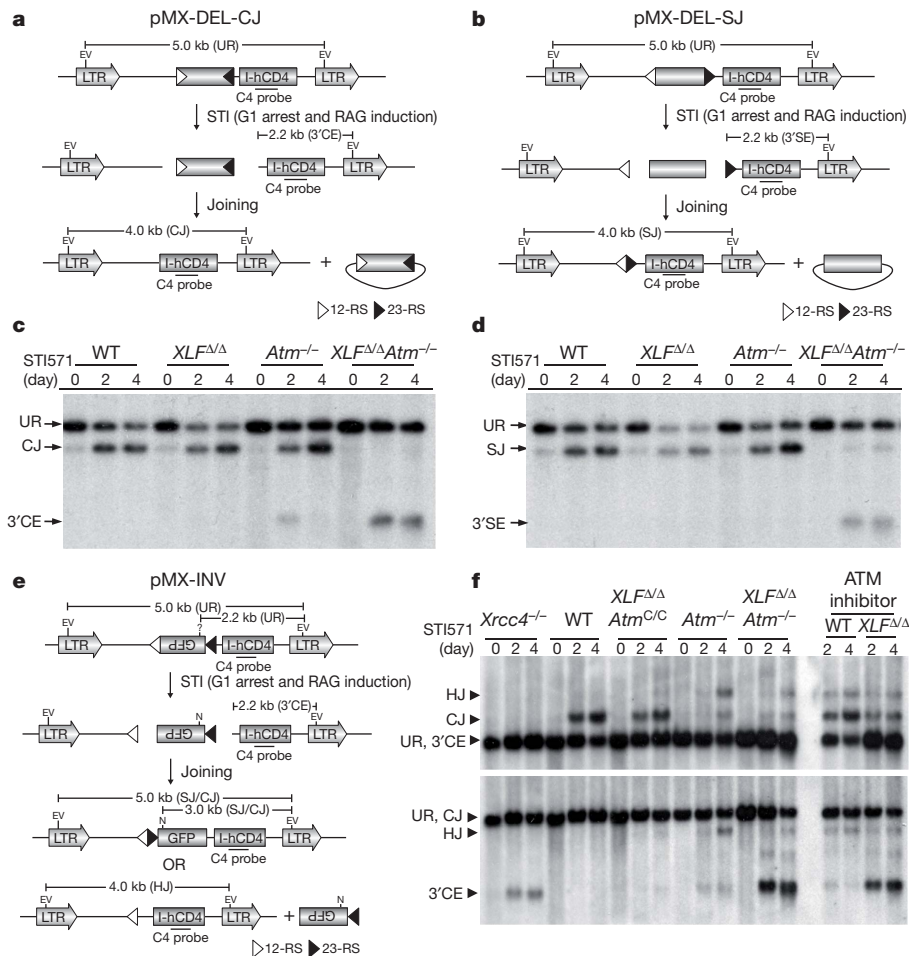


Figure 2 | ATM and XLF have redundant functions in chromosomal V(D)J recombination. **a, b, e**, Schematic of pMX-DEL-CJ (**a**), pMX-DEL-SJ (**b**) and pMX-INV (**e**) retroviral recombination substrates designed to assay CJ, SJ and inversive V(D)J recombination, respectively⁸. Diagrams indicate un-rearranged substrate (UR), coding/signal end (CE/SE) intermediates and coding/signal joints (CJ/SJ). The 12-recombination signal sequence (12-RS; open triangle), GFP coding sequence, 23-recombination signal sequence (23-RS; filled triangle), IRES-truncated hCD4 cDNA (I-hCD4) and LTRs are indicated. Positions of EcoRV (EV) sites, NcoI (N) sites and C4 probe (black bar) are shown. **c, d**, Southern blotting with C4 probe of EcoRV-digested DNA

functions, we inter-crossed *XLF Δ/Δ* mice that were heterozygous for an inactivating mutation of H2AX²² (*H2AX $^{F/F}$* ; *H2AX* also called *H2afx*). Notably, these crosses yielded no *XLF Δ/Δ H2AX $^{-/-}$* pups, with embryonic death of double homozygous mutants occurring before embryonic day 13.5 (Table 1). The finding that combined XLF and H2AX deficiency, but not combined XLF and ATM deficiency, is embryonically lethal might have several explanations. One is that the lethality results from ATM-independent S-phase functions of H2AX. Another is that impaired checkpoint functions associated with ATM deficiency rescue downstream effects of classical NHEJ deficiency that, otherwise, could result in embryonic death²³.

To determine whether XLF and H2AX have overlapping V(D)J recombination functions, we generated *XLF Δ/Δ* mice that were homozygous for a loxP-flanked H2AX allele (*H2AX $^{F/F}$*)^{21,22}. From these mice, we generated *v-abl* transformed *XLF Δ/Δ H2AX $^{F/F}$* pro-B lines containing an integrated single-copy inversional V(D)J recombination substrate (Fig. 4a, b and Supplementary Fig. 5). We then used Cre recombinase to generate matched sets of *XLF Δ/Δ H2AX $^{F/F}$* and *XLF Δ/Δ H2AX $^{-/-}$* lines, which were treated with STI571 and assayed for V(D)J recombination via GFP expression (Supplementary Fig. 5). In six matched sets, each with a different substrate integration, H2AX

deletion reduced, but did not eliminate, V(D)J recombination (Fig. 4a, b and Supplementary Fig. 5), indicating that XLF and H2AX also have overlapping V(D)J recombination activities, but not to the same extent as XLF and ATM. We also assayed for CJs, hybrid joins and unjoined CEs by Southern blotting in three matched sets of *XLF Δ/Δ H2AX $^{F/F}$* and *XLF Δ/Δ H2AX $^{-/-}$* pro-B lines (Fig. 4a, b and Supplementary Fig. 4b). *XLF Δ/Δ H2AX $^{F/F}$* lines behaved like wild-type or *XLF Δ/Δ* lines, as they robustly generated CJs but not hybrid joins or unjoined CEs (Fig. 4a, b and Supplementary Fig. 5). However, in accord with GFP assays, *XLF Δ/Δ H2AX $^{-/-}$* pro-B lines had substantially reduced CJs compared to *XLF Δ/Δ H2AX $^{F/F}$* parents, but did not show readily detectable unjoined CEs (Fig. 4a, b and Supplementary Fig. 5).

Recent studies found that H2AX protects unjoined coding ends from ATM kinase and CtIP-dependent resection²⁴. To test whether reduced V(D)J recombination in *XLF Δ/Δ H2AX $^{-/-}$* pro-B lines resulted from reduced RAG cleavage or reduced joining of RAG cleaved ends coupled with end resection, we performed V(D)J joining assays in *XLF Δ/Δ H2AX $^{F/F}$* and *XLF Δ/Δ H2AX $^{-/-}$* pro-B lines treated with an ATM kinase inhibitor (Fig. 4a, b and Supplementary Fig. 5). ATM inhibitor treatment of the *XLF Δ/Δ H2AX $^{F/F}$* lines reproduced the phenotype of *XLF Δ/Δ Atm $^{-/-}$* lines, including severely reduced CJs and the

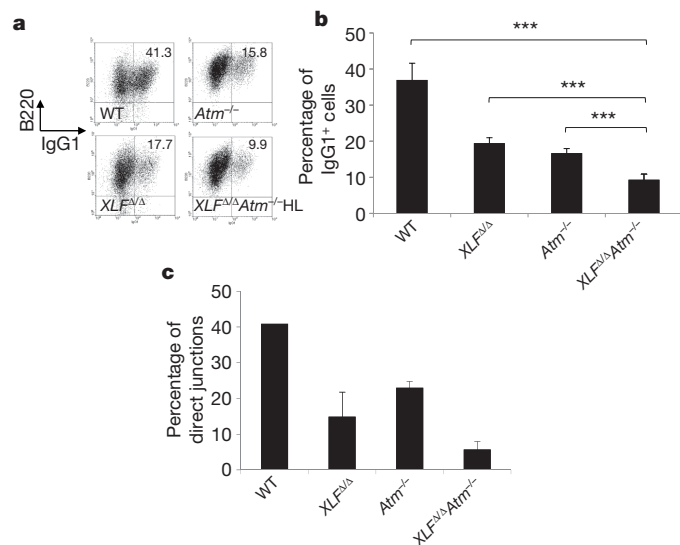


Figure 3 | ATM and XLF synergize in classical NHEJ during CSR.

a, Representative flow cytometric analysis of (at least three independent experiments for each genotype) purified CD43⁺ splenocytes from the indicated mice stained for surface B220 and surface IgG1 after a 4-day stimulation with anti-CD40 and IL-4. Additional experiments are in Supplementary Fig. 3. **b**, Summary of IgG1 CSR levels of purified CD43⁺ splenocytes after 4 days of anti-CD40 plus IL-4 stimulation. The y-axis shows the average percentage of IgG1⁺ cells determined from multiple experiments with cells from wild type ($n = 5$), *XLF*^{Δ/Δ} ($n = 5$), *Atm*^{-/-} ($n = 5$) and *XLF*^{Δ/Δ}*Atm*^{-/-} HL ($n = 8$) mice. Error bars show standard deviations. *** $P < 0.001$, based on student's *t*-test between indicated pairs. **c**, Percentage of direct junctions relative to direct plus microhomology (MH)-mediated junctions between Sμ and Sγ1 in anti-CD40-plus IL-4-stimulated B cells. Junctions were obtained from multiple independent experiments with *XLF*^{Δ/Δ} ($n = 4$), *Atm*^{-/-} ($n = 3$) and *XLF*^{Δ/Δ}*Atm*^{-/-} HL ($n = 4$) cells. See supplementary Fig. 4 for details.

accumulation of unjoined CE (Figs 2f, 4a, b and Supplementary Fig. 5). However, the ATM-inhibitor-treated *XLF*^{Δ/Δ}*H2AX*^{-/-} lines now yielded a clear band of unjoined CE associated with a CE smear below the band that is characteristic of aberrant end resection (Fig. 4a, b and Supplementary Fig. 5)²⁴. To examine this phenomenon further, we used a sensitive TdT end labelling assay²⁴, which revealed unjoined coding ends in STI571-treated *XLF*^{Δ/Δ}*H2AX*^{-/-} pro-B lines without ATM inhibitor treatment (Supplementary Fig. 6). These results demonstrate that XLF and H2AX have overlapping activities in V(D)J recombinational joining and also indicate that the unjoined V(D)J CE in *XLF*^{Δ/Δ}*H2AX*^{-/-} pro-B lines are largely resected in the absence of H2AX. Because we did not find complete restoration of the unjoined coding ends in ATM-kinase-inhibitor-treated *XLF*^{Δ/Δ}*H2AX*^{-/-} pro-B lines, as was observed in ATM-inhibitor-treated classical NHEJ-deficient cells that also are H2AX deficient²⁴, XLF might also have an overlapping function with H2AX in end protection.

We consistently observed a lower level of TdT-labelled CE in *XLF*^{Δ/Δ} versus *XLF*^{Δ/Δ}*H2AX*^{-/-} pro-B lines after STI571 induction in the presence of ATM inhibitor, even though the latter have substantially higher levels of unjoined CE (Fig. 4 and Supplementary Figs 5 and 6). This finding indicated that unjoined CE in ATM-inhibitor-treated *XLF*^{Δ/Δ} pro-B lines could be blocked from TdT activity. We used urea denaturing gel electrophoresis to test for a defect in opening

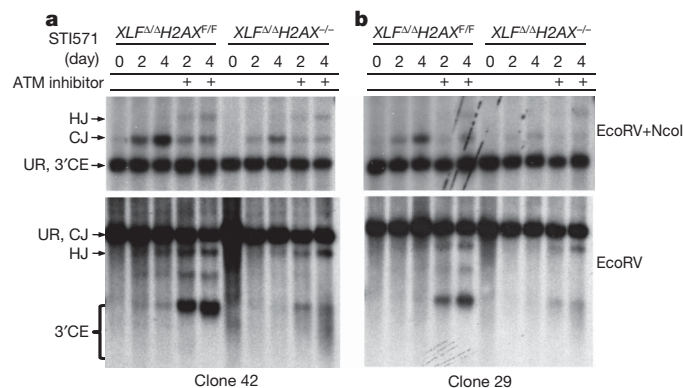


Figure 4 | H2AX and XLF have redundant functions. **a**, **b**, Southern blot analyses of rearrangement of a clonally integrated single-copy pMX-INV inversional V(D)J recombination substrate in *XLF*^{Δ/Δ}*H2AX*^{-/-} and derivative *H2AX* deleted *XLF*^{Δ/Δ}*H2AX*^{-/-} lines. **a**, Clone 42. **b**, Clone 29. Top: DNA was digested with EcoRV and NcoI and probed with C4 probe. Bottom: DNA was digested with EcoRV and probed with C4 probe (see Fig. 2e and legend for additional details). Analysis of an independent line is presented in Supplementary Fig. 5.

coding-end hairpins and found that ATM-inhibitor-treated *XLF*^{Δ/Δ} pro-B lines, like *Artemis*^{-/-} but not *Xrcc4*^{-/-} lines, indeed accumulated unopened hairpin CE (Supplementary Fig. 7). Given the overlapping functions of the ATM kinase and DNA-PK in CSR²⁵, plus the role of DNA-PK in activating Artemis to cleave hairpin coding ends², it seemed possible that dual deficiency for ATM and XLF leads to a DNA-PK defect. However, measurements of ionizing-radiation-induced phosphorylation of H2AX and KAP-1—both ATM and DNA-PK substrates²⁵—in the presence or absence of DNA-PK inhibitors revealed DNA-PK kinase activity to be as active in *XLF*^{Δ/Δ}*Atm*^{-/-} cells as in wild-type or *Atm*^{-/-} cells (Supplementary Fig. 8). However, we cannot rule out the possibility that *XLF*^{Δ/Δ}*Atm*^{-/-} cells have a specific defect in DNA-PK activity in the context of V(D)J recombination joining that overlaps with ATM/XLF functions or that a DNA-PK defect in this context could represent a more general defect, for example in the broader recruitment of classical NHEJ factors. In the latter context, we note that the V(D)J recombination defect in *XLF*^{Δ/Δ}*Atm*^{-/-} pro-B lines is not just in hairpin opening, as we also observe a severe defect in recombination signal sequence joining in these cells (Fig. 2).

ATM, XLF and H2AX previously have been found to have, at most, modest roles in V(D)J recombination and, by extension, classical NHEJ^{5,7-9}. Surprisingly, we now find that dual deficiency for XLF and ATM has an impact on V(D)J recombination in progenitor lymphocytes and IgH CSR in mature B cells similarly to deficiency for bona fide classical NHEJ factors. We conclude that XLF-deficient cells require ATM, and that ATM-deficient cells require XLF, to carry out classical NHEJ but not A-EJ. These findings further indicate that XLF-deficient cells provide a novel system for elucidating major, previously unappreciated roles of ATM and ATM substrates in classical NHEJ and vice versa. Our findings already suggest that ATM, via phosphorylation of its substrates, and XLF share overlapping functions primarily in the context of chromosomal classical NHEJ. We have also shown that the V(D)J recombination defects in XLF/ATM- and XLF/H2AX-deficient pro-B lines are not identical either in extent or in outcome. Thus, it remains possible that additional ATM substrates, besides

Table 1 | H2AX and XLF have redundant functions in murine embryonic development

| Stage | <i>H2AX</i> ^{+/+} <i>XLF</i> ^{Δ/Δ} | <i>H2AX</i> ^{+/+} <i>XLF</i> ^{Δ/Δ} | <i>H2AX</i> ^{-/-} <i>XLF</i> ^{Δ/Δ} | Absorbed | Total |
|------------------|--|--|--|----------|-------|
| Birth | 14 | 46 | 0 | 0 | 60 |
| Birth (expected) | 15 | 30 | 15 | — | — |
| E13.5 | 8 | 28 | 0 | 8 | 44 |
| E13.5 (expected) | 11 | 22 | 11 | — | — |

The indicated genotypes were obtained by inter-crossing *H2AX*^{+/+}*XLF*^{Δ/Δ} mice as described^{5,12,22}. H2AX deficiency does not cause embryonic lethality^{22,27}.

H2AX, may also overlap functionally with XLF. XLF might directly influence the same processes as ATM or H2AX, including end processing and end resection, respectively. Alternatively, overlapping functions might be mediated indirectly through distinct processes. For example, XLF may influence reaction kinetics by classical NHEJ factor recruitment²⁶, whereas ATM and ATM substrates seem to tether chromosomal ends for joining^{7,8}—two distinct functions that theoretically could be redundant with respect to effects on overall joining activities.

METHODS SUMMARY

Mice. *XLF*^{+/-}, *Atm*^{+/-}, *Atm*^{+/-C}, *H2AX*^{+/-} and *H2AX*^{+/-F} and 'HL' mice have been described previously^{5,12,21,22}. All HL mice were heterozygous for both *IgH* and *IgL* knockin alleles.

Chromosomal V(D)J recombination assays. V(D)J recombination with an integrated substrate was carried out as described⁸. Briefly, *v-abl* transformed pro-B-cell lines were isolated from various mouse lines that harboured an *Eμ-Bcl-2* transgene. For XRCC4-deficient *v-abl* transformants, the *Eμ-Bcl-2* transgene was introduced after establishment of the line⁵. The pro-B lines were infected with the pMX-INV, pMX-DelCJ or pMX-DelSJ retroviral vector and assayed for V(D)J recombination as described^{5,8}. ATM inhibitor Ku55933 (Cat.No.118500 from EMD Biosciences) was used as a final concentration of 15 μM as described⁸.

Lymphocyte development and class switch recombination. Lymphocyte populations were analysed by flow cytometry as described⁵. Isolation and activation of splenic B cells and flow cytometric assays were as described¹³. Sp-Sy1 junctions were isolated from day 4 anti-CD40 plus IL-4 stimulated B cells, cloned and sequenced as described¹³.

Received 17 June; accepted 21 October 2010.

Published online 15 December 2010.

1. Lieber, M. R. The mechanism of human nonhomologous DNA end joining. *J. Biol. Chem.* **283**, 1–5 (2008).
2. Rooney, S., Chaudhuri, J. & Alt, F. W. The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol. Rev.* **200**, 115–131 (2004).
3. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **124**, 301–313 (2006).
4. Buck, D. *et al.* Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**, 287–299 (2006).
5. Li, G. *et al.* Lymphocyte-specific compensation for XLF/Cernunnos end-joining functions in V(D)J recombination. *Mol. Cell* **31**, 631–640 (2008).
6. Zha, S., Alt, F. W., Cheng, H. L., Brush, J. W. & Li, G. Defective DNA repair and increased genomic instability in Cernunnos-XLF-deficient murine ES cells. *Proc. Natl Acad. Sci. USA* **104**, 4518–4523 (2007).
7. Bassing, C. H. & Alt, F. W. The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst.)* **3**, 781–796 (2004).
8. Bredemeyer, A. L. *et al.* ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466–470 (2006).
9. Yin, B. *et al.* Histone H2AX stabilizes broken DNA strands to suppress chromosome breaks and translocations during V(D)J recombination. *J. Exp. Med.* **206**, 2625–2639 (2009).
10. Bassing, C. H., Swat, W. & Alt, F. W. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109** (Suppl.), S45–S55 (2002).

11. Bredemeyer, A. L. *et al.* DNA double-strand breaks activate a multi-functional genetic program in developing lymphocytes. *Nature* **456**, 819–823 (2008).
12. Zha, S., Sekiguchi, J., Brush, J. W., Bassing, C. H. & Alt, F. W. Complementary functions of ATM and H2AX in development and suppression of genomic instability. *Proc. Natl Acad. Sci. USA* **105**, 9302–9306 (2008).
13. Yan, C. T. *et al.* IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
14. Gellert, M. Molecular analysis of V(D)J recombination. *Annu. Rev. Genet.* **26**, 425–446 (1992).
15. Corneo, B. *et al.* Rag mutations reveal robust alternative end joining. *Nature* **449**, 483–486 (2007).
16. Chaudhuri, J. *et al.* Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv. Immunol.* **94**, 157–214 (2007).
17. Franco, S. *et al.* H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* **21**, 201–214 (2006).
18. Ramiro, A. R. *et al.* Role of genomic instability and p53 in AID-induced c-myc-IgH translocations. *Nature* **440**, 105–109 (2006).
19. Lumsden, J. M. *et al.* Immunoglobulin class switch recombination is impaired in *Atm*-deficient mice. *J. Exp. Med.* **200**, 1111–1121 (2004).
20. Reina-San-Martin, B., Chen, H. T., Nussenzweig, A. & Nussenzweig, M. C. ATM is required for efficient recombination between immunoglobulin switch regions. *J. Exp. Med.* **200**, 1103–1110 (2004).
21. Bassing, C. H. *et al.* Increased ionizing radiation sensitivity and genomic instability in the absence of histone H2AX. *Proc. Natl Acad. Sci. USA* **99**, 8173–8178 (2002).
22. Bassing, C. H. *et al.* Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. *Cell* **114**, 359–370 (2003).
23. Sekiguchi, J. *et al.* Genetic interactions between ATM and the nonhomologous end-joining factors in genomic stability and development. *Proc. Natl Acad. Sci. USA* **98**, 3243–3248 (2001).
24. Helmink, B. A. *et al.* H2AX prevents CtIP-mediated DNA end resection and aberrant repair in G1-phase lymphocytes. *Nature* doi:10.1038/nature09585 (this issue).
25. Callén, E. *et al.* Essential role for DNA-PKcs in DNA double-strand break repair and apoptosis in ATM-deficient lymphocytes. *Mol. Cell* **34**, 285–297 (2009).
26. Yano, K. *et al.* Ku recruits XLF to DNA double-strand breaks. *EMBO Rep.* **9**, 91–96 (2008).
27. Celeste, A. *et al.* Genomic instability in mice lacking histone H2AX. *Science* **296**, 922–927 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Fujiwara and P. Huang for technical support. We thank B. Sleckman for advice, reagents and for critical review of this manuscript. This work is supported by NIH grant AI076210 and AI020047 to F.W.A. F.W.A. is an investigator of the Howard Hughes Medical Institute. S.Z. was a fellow, then senior fellow of Leukemia and Lymphomas Society of America and a St Baldrick Scholar. C.G. and Y.Z. are fellows of Cancer Research Institute. C.B. receives support from the pre-doctoral training program of Cancer Research Institute. D.R.W. is supported by a career development award from AAI/GlaxoSmithKline and by NIH training grant AI007376.

Author Contributions S.Z., C.G. and F.W.A. designed experiments and wrote the paper. S.Z., C.G., C.B., V.O., H.-L.C., Y.Z., D.R.W., G.Y., H.P., P.H.G. and R.L.D. performed experiments.

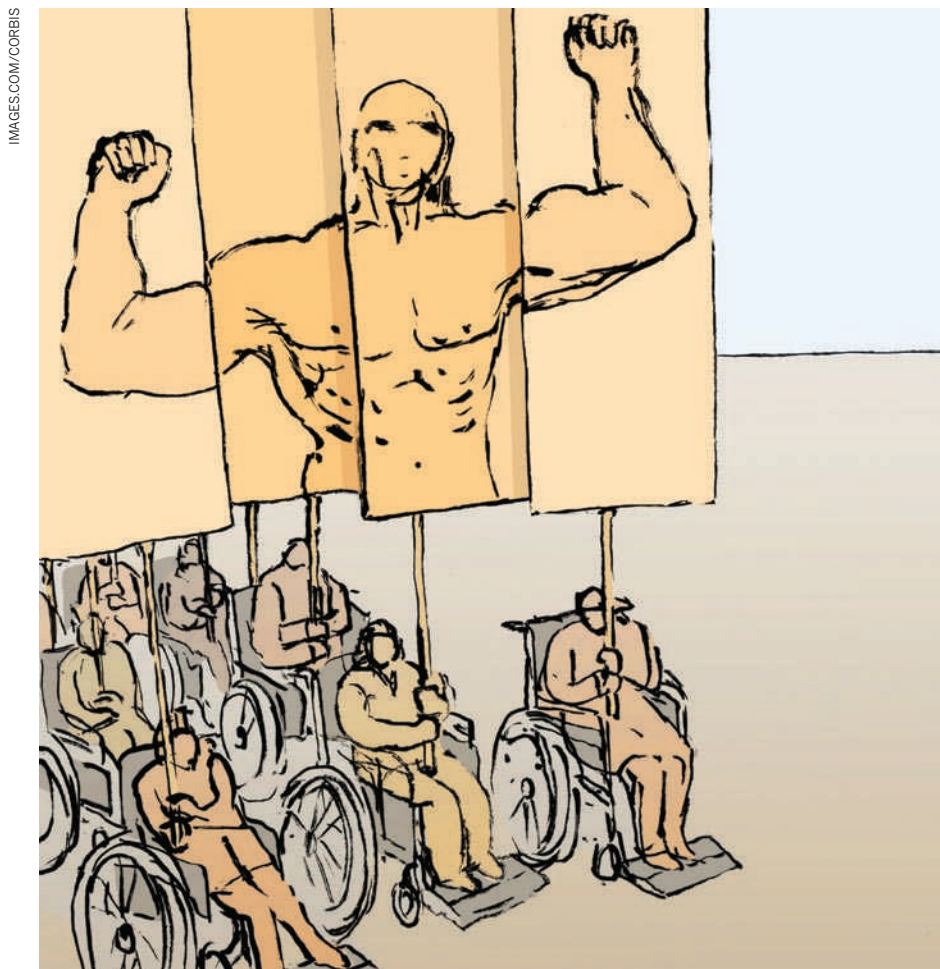
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.W.A. (alt@enders.tch.harvard.edu).

CAREERS

TURNING POINT A switch from chemistry to biology led to research success **p.257**

NATURE NETWORK For more discussion of careers issues, visit go.nature.com/lm1x4t

NATUREJOBS For the latest career listings and advice www.naturejobs.com



IMAGES.COM/CORBIS

EQUALITY

The fight for access

Scientists with disabilities seek ways to level the playing field.

BY VIRGINIA GEWIN

Jesse Leaman is, in some ways, typical of the many young, promising scientists struggling to excel in a tough funding environment. After receiving a PhD in astrophysics from the University of California, Berkeley, and completing a postdoc last summer at NASA Ames

Research Center in Moffett Field, California, he is now looking for work. Not so typical, however, are the obstacles that Leaman faces on a daily basis — he is paralysed from the neck down as a result of an accident at the age of 18.

Leaman chalks up most of his employment travails to being in a highly competitive field during an economic downturn. “I’m struggling

along with everyone else,” he says. But he is concerned that his disability will compound an already challenging job search. Travelling, for example, requires him to arrange for nursing care, so it isn’t easy to spend, say, a week in Europe to interview for positions. And Leaman can’t earn extra cash through the typical routes while he looks for work. “I can’t work as a waiter while I wait for grant proposals to go through,” he says.

Although Leaman and others with disabilities face complex challenges, the job landscape in the United States has largely improved since the 1990 signing of the Americans with Disabilities Act, which prevents discrimination on the basis of disability. And in the United Kingdom, the Equality Act, which took effect last October, provides that country’s most comprehensive anti-discrimination measures ever, including provisions against indirect discrimination — for example, any employment condition that particularly disadvantages disabled people. In the United States and parts of Europe, shifting societal attitudes, advances in technologies to control laboratory equipment remotely and funding schemes aimed at boosting training opportunities for people with disabilities make scientific careers more accessible than ever before. But advocates for the disabled say that the ability to honestly assess and communicate abilities as well as disabilities will be most helpful to the careers of Leaman and others in similar situations.

SHIFTING ATTITUDES

Disabilities have a real impact on a person’s capacity to get a job. People with disabilities — physical, mental or emotional conditions that impair activity — consistently have a higher unemployment rate than the general population. For November 2010, the US Bureau of Labor Statistics reported an unemployment rate of 14.5% among civilians aged 16 and older who had a disability, compared with 9.1% among those who did not have a disability. Data on employment rates for disabled people are scarce in Europe, but labour-force surveys and some data from the *European Union Statistics on Income and Living Conditions* database for 2008 suggest that well below 75% of people with disabilities are in employment, says the Academic Network of Disability Experts, which aids the European Commission in developing disability policy. The best employment rate for people with disabilities hovers at a meagre 50–55% in countries such as Sweden and Germany, whereas Romania, ►

► Hungary and Poland have a dismal rate of 30%. Low rates of employment may be linked to any number of factors, including a lack of training, opportunities or confidence.

Simply getting into graduate school is the first hurdle. According to the US National Science Foundation (NSF), 10% of undergraduates entering science, technology, engineering or maths (STEM) fields have disabilities. Yet typically only 1% of people who receive STEM doctorates have disabilities — roughly 355 people in 2008. In 2004, the number hit a low of 284, part of a three-year decline that spurred NSF efforts to foster more opportunities, says Mark Leddy, director of the NSF's programme for research in disabilities education based in Arlington, Virginia.

Increases in the number of people with disabilities applying for STEM degrees and medical school may also reflect the fact that the term 'disability' has come to comprise a broader range of disorders than it once did. The traditional definition referred to someone who used a wheelchair or was visually or hearing impaired, but now it often also includes people with learning disabilities, attention-deficit disorders and psychiatric conditions. In Britain, for example, people with dyslexia and Asperger's syndrome submit among the largest number of preliminary enquiries for the government's Disabled Student Allowance (DSA), which was established in 1993 to help universities provide assistance such as sign-language interpreters, note-takers or assistive computer software.

BEYOND FUNDING CHALLENGES

In the United States, a growing number of federal programmes offers supplemental funding or specific schemes to help students secure employment at research institutions. For example, the National Institutes of Health (NIH) in Bethesda, Maryland, offers grant supplements to promote diversity, and provides PhD and MD fellowship awards for students with disabilities. The NSF offers facilitation awards for scientists and engineers with disabilities, to cover the costs of special accommodation or equipment needed to do research. It also funds ten university-based alliances, usually across multiple campuses throughout the country, that offer career support including mentoring and internships for people with disabilities who aspire to a scientific career. Other NSF-funded programmes, such as DO-IT (Disabilities, Opportunities, Internetworking and Technology), based at the University of Washington in Seattle, offer online mentoring and support to identify adaptive technologies, software, and online resources that can improve access to academic programmes.

In Britain, the outlook for postgraduate students with disabilities is less positive. One reason is that the DSA available to postgraduates is roughly half of that available to undergraduates. As a result, supplemental schemes are in high demand among postgraduates, says Paul Alexander, chief executive of the Snowdon Awards Scheme, a private funding body based in Southwater that was created in 1981 to facilitate disabled students' pursuit of higher education. The scheme's maximum award is only £2,500 (US\$3,900), but Alexander says that people with disabilities seeking postgraduate education make up at least 50% of the Snowdon scheme's roughly 100 awards each year. "Supplemental funding

— go largely untapped among the disabled, although they are eligible for supplemental awards to cover travel- or care-related costs. Just 213 students with special needs — out of a total of 198,523 Erasmus students Europe-wide — received extra funding during 2008–09. In Britain, only 9 out of 11,724 Erasmus students applied for the supplemental awards in 2010, according to the British Council, which oversees the scheme in the United Kingdom. As Leaman notes, travel is difficult for some with physical disabilities, but Priestley says that the apparent lack of take-up of mobility schemes may also reflect how the inconsistent patchwork of legislation covering support services and employment in European countries can make such a move daunting to people with disabilities.

There are some hurdles that targeted schemes can't help with. "Getting started as a young investigator is the hardest part for scientists with physical disabilities, given attitudinal as well as physical barriers that must be overcome," says Brad Duerstock, a biomedical researcher at Purdue University in West Lafayette, Indiana, who uses a wheelchair. Duerstock says that he was lucky to do his PhD at the Center for Paralysis Research at Purdue with Richard Borgens, a neuroscientist. "He knew my limitations and was very open about working with me to find out what kind of research I could do," says Duerstock. Still, he says, he struggled to hit training milestones as quickly as people without disabilities.

Duerstock plans to open an Institute for Accessible Science at Purdue this year — launched with a US\$2-million NIH 'Pathfinder' award that he received last October. The awards aim to increase the diversity of the scientific workforce. Duerstock's reconfigurable lab space will allow him to help at least six young disabled scientists work out how best to modify laboratory equipment so that they can conduct science autonomously. "Our main mission is to help these students be as independent as possible — that's what they need," he says.

But technology alone won't prepare disabled scientists for the job market. Internships are key: they offer access to a potential employer while letting employers assess abilities without making long-term commitments. To be frank, says Sheryl Burgstahler, director of DO-IT, some employers may be nervous about hiring someone with a disability because, if things don't work out, anti-discrimination laws can make termination difficult. Internships, she says, are a way for both the employer and intern to have a positive experience — and have led several participants to permanent positions at Microsoft, for example.

The Entry Point! programme, funded by the



Mark Priestley of the University of Leeds, UK, pictured here with a colleague, critiques European policy on disability.

is really uncommon throughout Europe, which can lead to all sorts of problems," says Mark Priestley, professor of disability policy at the University of Leeds, UK. For example, he says, students with disabilities who take longer than expected to complete their PhDs as a result of chronic health conditions may find that their grants don't adequately fund the extra time needed. "We've had a few issues with students needing extensions, yet the funders were not flexible," he says. In those cases, the funding simply runs out.

Other opportunities, such as the Erasmus scheme — designed to facilitate the movement of student researchers across Europe

American Association for the Advancement of Science in Washington DC, facilitates internships to NASA research centres, national research laboratories and corporations such as IBM. “We’re not only preparing them for the internship, we’re building a support network to encourage them to go on to graduate school,” says Laureen Summers, who helps to place Entry Point! interns at NASA facilities. Similarly, the Career Opportunities for Students with Disabilities (COSD) programme, based at the University of Tennessee in Knoxville but covering more than 600 universities, facilitates internships at companies such as Dow Chemical, Microsoft and CISCO. “At Dow, we’re always

“The essence of science is being able to think creatively, and sometimes the most creative insights come from people who think about the world differently.”

looking for innovative people. Persons with disabilities have certainly dealt with obstacles that require creative approaches,” says Shawn Loachridge, the sourcing and global university relations manager at Dow’s company workforce planning office in Midland,

Michigan. Students with disabilities represent an untapped market for employers increasingly eager to boost workplace diversity, says Alan Muir, director of COSD.

Although these programmes help students with disabilities to gain employment opportunities, successful job bids require frank self-assessment. “It takes a realistic view of one’s own strengths and weaknesses for a person to become an effective self-advocate,” says Carla Romney, chair of the undergraduate science and engineering programme at Boston University in Massachusetts, who has spent the past decade creating programmes to help students with disabilities to engage in STEM. Romney cautions students against overestimating their physical abilities. At the same time, they should be frank about what they need to succeed, and ask for it. She says that although attitudes are changing, many students are still reluctant to disclose their disabilities on application forms for fear that they won’t be admitted — which, she says, is a disservice to themselves. Burgstahler and Romney say that it is imperative that students learn the communication skills necessary not only to become their own advocates, but also to become good scientists. “The essence of science,” says Romney, “is being able to think creatively, and sometimes the most creative insights come from people who think about the world differently.” ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

TURNING POINT William Ja

William Ja, assistant professor of metabolism and ageing at the Scripps Research Institute in Jupiter, Florida, is getting noticed — and funded — for his cross-disciplinary research in protein chemistry and fruitfly genetics.

Did you intentionally leap from a PhD in chemistry into a postdoc in biology?

Yes and no. I majored in chemistry at the University of California, Berkeley, as an undergraduate, but always knew I wanted to do something biological. I just didn’t know what. While getting my chemistry PhD at the California Institute of Technology in Pasadena, I wanted to bridge the gap between chemistry and biology, so I switched to a postdoc in pure biology. Then I got the chance to work with Seymour Benzer, a pioneer in molecular biology known for unravelling the ties between genes, behaviour and longevity. I loved it, but realized how difficult it is to stay at the forefront of two disciplines.

How did you come to work with Benzer?

My PhD focused on developing techniques to help peptides bind to protein targets. Benzer’s work on the long-lived fruitfly mutant dubbed Methuselah was well known on campus. I wondered whether I could use my technique to develop peptides that could bind to Methuselah, to see whether they could affect ageing in flies. At the time, the Glenn Foundation for Medical Research in Carpinteria, California, had a scholarship programme that aimed to give PhD students a sabbatical-like experience. I applied, hoping to join Benzer’s lab to do work on ageing *in vivo*. After that, Benzer invited me back as a postdoc.

What did you learn from that experience?

We found that certain peptides, when over-expressed, act as longevity drugs that can extend the lifespan of the fly. I found that I loved working on the organismal level. As a chemist, I had never worked with things that walk. I got interested in fly behaviour, courtship, memory and learning. And, because I had no biology experience, I asked naive questions — for example, I examined flies’ eating and drinking habits.

Do you have a career strategy?

I enjoy working on questions that are fun and interesting and that no one else is exploring. One example is measuring how much flies eat, which is difficult in such a small organism. Another is whether fruitflies get thirsty, which questions a 100-year tradition of rearing them in the lab without free water. We found that diet and thirst can affect a variety of fruitfly



J. MCENTEE/SCRIPPS

phenotypes. My papers are often the result of ideas thrown out at lunch or over drinks — I’ve had a 10–20% success rate following up on questions conceived in a bar. That is how two of my best papers — the questions I mention above — began.

How were the diet and thirst findings received?

Some people doing fruitfly research get upset when I present my results, because my findings might cast doubt on their own work. In those instances, I’ve encouraged authors to test whether thirst makes a difference and publish their results — that is the only way that we’ll know if this is a significant problem. Debates are part of science, but I would say that the politics of science has been the biggest surprise to me as I’ve moved into a more independent stage of my career. I didn’t realize, as a post-graduate student, how shielded I was from politics. I’ve learned that being a good scientist doesn’t mean you are always right.

As a new assistant professor in a bleak economy, where are you looking for money?

There is a lot of stress in this economy about getting grant money, but I do science best when I’m working on diverse projects. The Glenn Foundation surprised me with an unsolicited US\$60,000 award last November to continue my research into ageing. And in October, I received funding to explore non-surgical sterilization methods for dogs and cats from the Found Animals Foundation in Los Angeles, California, a non-profit organization dedicated to minimizing euthanasia at animal shelters. It had a prize for creating a single-dose sterilant for cats and dogs, and I wondered about using a cytotoxin to attack cells important for reproduction. I wrote a proposal capitalizing on my protein-chemistry and molecular-biology experience and asked for \$200,000. They agreed. ■

INTERVIEW BY VIRGINIA GEWIN

American Association for the Advancement of Science in Washington DC, facilitates internships to NASA research centres, national research laboratories and corporations such as IBM. “We’re not only preparing them for the internship, we’re building a support network to encourage them to go on to graduate school,” says Laureen Summers, who helps to place Entry Point! interns at NASA facilities. Similarly, the Career Opportunities for Students with Disabilities (COSD) programme, based at the University of Tennessee in Knoxville but covering more than 600 universities, facilitates internships at companies such as Dow Chemical, Microsoft and CISCO. “At Dow, we’re always

“The essence of science is being able to think creatively, and sometimes the most creative insights come from people who think about the world differently.”

looking for innovative people. Persons with disabilities have certainly dealt with obstacles that require creative approaches,” says Shawn Loachridge, the sourcing and global university relations manager at Dow’s company workforce planning office in Midland,

Michigan. Students with disabilities represent an untapped market for employers increasingly eager to boost workplace diversity, says Alan Muir, director of COSD. Although these programmes help students with disabilities to gain employment opportunities, successful job bids require frank self-assessment. “It takes a realistic view of one’s own strengths and weaknesses for a person to become an effective self-advocate,” says Carla Romney, chair of the undergraduate science and engineering programme at Boston University in Massachusetts, who has spent the past decade creating programmes to help students with disabilities to engage in STEM. Romney cautions students against overestimating their physical abilities. At the same time, they should be frank about what they need to succeed, and ask for it. She says that although attitudes are changing, many students are still reluctant to disclose their disabilities on application forms for fear that they won’t be admitted — which, she says, is a disservice to themselves. Burgstahler and Romney say that it is imperative that students learn the communication skills necessary not only to become their own advocates, but also to become good scientists. “The essence of science,” says Romney, “is being able to think creatively, and sometimes the most creative insights come from people who think about the world differently.” ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

TURNING POINT William Ja

William Ja, assistant professor of metabolism and ageing at the Scripps Research Institute in Jupiter, Florida, is getting noticed — and funded — for his cross-disciplinary research in protein chemistry and fruitfly genetics.

Did you intentionally leap from a PhD in chemistry into a postdoc in biology?

Yes and no. I majored in chemistry at the University of California, Berkeley, as an undergraduate, but always knew I wanted to do something biological. I just didn’t know what. While getting my chemistry PhD at the California Institute of Technology in Pasadena, I wanted to bridge the gap between chemistry and biology, so I switched to a postdoc in pure biology. Then I got the chance to work with Seymour Benzer, a pioneer in molecular biology known for unravelling the ties between genes, behaviour and longevity. I loved it, but realized how difficult it is to stay at the forefront of two disciplines.

How did you come to work with Benzer?

My PhD focused on developing techniques to help peptides bind to protein targets. Benzer’s work on the long-lived fruitfly mutant dubbed Methuselah was well known on campus. I wondered whether I could use my technique to develop peptides that could bind to Methuselah, to see whether they could affect ageing in flies. At the time, the Glenn Foundation for Medical Research in Carpinteria, California, had a scholarship programme that aimed to give PhD students a sabbatical-like experience. I applied, hoping to join Benzer’s lab to do work on ageing *in vivo*. After that, Benzer invited me back as a postdoc.

What did you learn from that experience?

We found that certain peptides, when over-expressed, act as longevity drugs that can extend the lifespan of the fly. I found that I loved working on the organismal level. As a chemist, I had never worked with things that walk. I got interested in fly behaviour, courtship, memory and learning. And, because I had no biology experience, I asked naive questions — for example, I examined flies’ eating and drinking habits.

Do you have a career strategy?

I enjoy working on questions that are fun and interesting and that no one else is exploring. One example is measuring how much flies eat, which is difficult in such a small organism. Another is whether fruitflies get thirsty, which questions a 100-year tradition of rearing them in the lab without free water. We found that diet and thirst can affect a variety of fruitfly



J. MCENTEE/SCRIPPS

phenotypes. My papers are often the result of ideas thrown out at lunch or over drinks — I’ve had a 10–20% success rate following up on questions conceived in a bar. That is how two of my best papers — the questions I mention above — began.

How were the diet and thirst findings received?

Some people doing fruitfly research get upset when I present my results, because my findings might cast doubt on their own work. In those instances, I’ve encouraged authors to test whether thirst makes a difference and publish their results — that is the only way that we’ll know if this is a significant problem. Debates are part of science, but I would say that the politics of science has been the biggest surprise to me as I’ve moved into a more independent stage of my career. I didn’t realize, as a post-graduate student, how shielded I was from politics. I’ve learned that being a good scientist doesn’t mean you are always right.

As a new assistant professor in a bleak economy, where are you looking for money?

There is a lot of stress in this economy about getting grant money, but I do science best when I’m working on diverse projects. The Glenn Foundation surprised me with an unsolicited US\$60,000 award last November to continue my research into ageing. And in October, I received funding to explore non-surgical sterilization methods for dogs and cats from the Found Animals Foundation in Los Angeles, California, a non-profit organization dedicated to minimizing euthanasia at animal shelters. It had a prize for creating a single-dose sterilant for cats and dogs, and I wondered about using a cytotoxin to attack cells important for reproduction. I wrote a proposal capitalizing on my protein-chemistry and molecular-biology experience and asked for \$200,000. They agreed. ■

INTERVIEW BY VIRGINIA GEWIN

THE PERFECT EGG

Domestic bliss?

BY TANIA HERSHMAN

He looks up and catches its eye. Eye? Silly! Visual circuitry. Optical sensors. But he's sure, he's sure it looked right at him. He eats his perfectly boiled egg. Can't stop himself from saying: "Thank you, this is just right," and swears he sees pleasure, just a hint, on its flawless face. Then it turns and begins to load the dishwasher. He dunks his toast into the runny yolk and tries not to dwell on it.

When he finishes, he gets up and puts his plate, knife and spoon into the sink. It is standing there, waiting.

"Please clean out the fridge, including the ice trays," he says. "They need defrosting." It nods. Is there a smile? I'm going mad, he thinks. He puts on his coat and leaves.

In the park he watches more of them sitting on benches, watching their charges in the playground. He's struck by what they *don't*. Don't fidget, scratch or mess with their hair. Don't turn their heads, chat with one another, read magazines or talk on mobile phones. They are absolutely still, completely focused. Just *there*.

He is tempted to run up and grab a child off the swings, just reach around its waist and pull the small body out, shrieking.

Just to see.

Just to know.

That night, he watches television while it irons in a corner of the living room. He is distracted from the sitcom that he won't admit he waits for each week by the smell of steaming fabric, the handkerchiefs he's had for 40 years or more, always neatly pressed. Worn a little, torn, but clean and wrinkle-free.

He stands up and, over by the ironing board, makes a big show of unzipping his fly.

No stirring. Not a flicker. It stops ironing and waits for further instructions.

He takes the trousers off, one leg and then the other, wobbling slightly as he tries to keep his dignity. He hands them over.

"Please do these too," he says, and sits back on the sofa in his underwear. He starts to laugh as, on the screen, the wife comes home and shouts at the useless husband.

Next morning, after another perfect egg with toast, he says: "Come with me." It walks behind him to the hall.

He opens the door to the cupboard underneath the stairs.

"Please go inside," he says, and it obeys. He shuts the door and goes upstairs to his study where for several hours in his head are words like *blackness, suffocation, boredom*.

He switches on the computer and writes a long e-mail to the woman who used to be his wife, rambling and without punctuation.



He says things he wishes he'd said in life, or in that life, at least. At first he calls it poetry and then he sees it's not. He deletes it and goes back down.

He walks about in the kitchen and from kitchen to living room, living room to downstairs bathroom. Then he stands in the hall, listening. He opens the cupboard door. Dark, no movement at all. It has no lights on. Oh my god, he thinks.

"Are you ...?" he says.

► **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

It whirs quickly out of Sleep mode.

"Please, come out," he says. It glides past him, nothing in its eyes or on its face. He has a sensation in his sinuses, unpleasant, unwelcome. He boils the kettle, leaves the full mug of tea on the counter, gets his coat and leaves.

In the park, he watches them again. Are they watching him watching them watching? He ambles over to the swings and puts a hand out, leaning on the rail as small girls giggle and try to touch the sky. No one moves or does anything. No one even looks in his direction.

How fast could they run if ...?

Would it be just the one who'd tackle him to the playground floor? Or all of them, some sort of instantaneous communication rousing them to action?

After a few minutes, the screams and creaking of the swings gives him shooting pains through his skull. He heads for home.

He eats dinner, listening to the radio, the evening news. He finishes, puts the plate in the sink, then he says: "Please come with me." And leads it upstairs. In the bedroom he instructs it to sit in the armchair in the corner. He puts on his pyjamas with some coyness, a wardrobe door shielding him. Then he gets into bed and pulls the covers tight around himself.

"Please watch," he tells it. "Just keep an eye. Make sure that nothing ... I mean, no sleeping."

He switches off his bedside light and can see a faint green glow coming from the armchair. He lies with his eyes open for a few moments and then he falls asleep.

In the morning, refreshed, he eats his perfect egg.

"Thank you," he says, and puts his plate, knife and fork into the sink. "Please do the carpets today," he tells it, and heads towards the stairs. ■

Tania Hershman's book *The White Road and Other Stories* (commended, Orange Award for New Writers) was included in New Scientist's Best Books of 2008. Tania is writer-in-residence at the University of Bristol's science faculty.

JACEY

Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*

Robin Hopkins¹ & Mark D. Rausher¹

Species formation generates biological diversity and occurs when traits evolve that prevent gene flow between populations. Discerning the number and distribution of genes underlying these traits and, in a few cases, identifying the genes involved, has greatly enhanced our understanding over the past 15 years of species formation (reviewed by Noor and Feder¹ and Wolf *et al.*²). However, this work has almost exclusively focused on traits that restrict gene flow between populations that have evolved as a by-product of genetic divergence between geographically isolated populations. By contrast, little is known about the characteristics of genes associated with reinforcement, the process by which natural selection directly favours restricted gene flow during the formation of species. Here we identify changes in two genes that appear to cause a flower colour change in *Phlox drummondii*, which previous work has shown contributes to reinforcement. Both changes involve *cis*-regulatory mutations to genes in the anthocyanin biosynthetic pathway (ABP). Because one change is recessive whereas the other is dominant, hybrid offspring produce an intermediate flower colour that is visited less by pollinators, and is presumably maladaptive. Thus genetic change selected to increase prezygotic isolation also appears to result in increased postzygotic isolation.

Natural selection can directly favour species formation through a process termed reinforcement. If two incipient species experience secondary contact and produce maladaptive hybrids, selection favours decreased gene flow and increased reproductive isolation between them^{3–5}. Reinforcement can be recognized by a resulting pattern known as reproductive character displacement: reproductive isolation is greater in sympatry than in allopatry⁴. Although the occurrence of reinforcement was historically controversial^{6,7}, empirical studies have documented reinforcing selection and reproductive character displacement in birds, insects, amphibians, plants and mammals, suggesting that it is a common step in the process of species formation (reviewed by Ortiz-Barrientos *et al.*⁸ and Pfennig and Pfennig⁹). Despite this work, little is known about genetic changes associated with reinforcement (but see studies of quantitative trait loci by Ortiz-Barrientos *et al.*¹⁰ and Saether *et al.*¹¹).

Divergence of floral colour in *P. drummondii* constitutes one of the best-documented cases of reinforcement in plants¹² and exhibits the classic pattern of reproductive character displacement. *P. drummondii* and the closely related *P. cuspidata* produce similar light-blue flowers throughout the allopatric parts of their ranges. However, in the area of sympatry, *P. drummondii* has dark-red flowers, representing the only natural evolution of red flowers in the *Phlox* clade¹². Both species of *Phlox* and colours of *P. drummondii* are pollinated by the same array of species of Lepidoptera¹². Hybrids between these two species are formed at rates as high as 11% in the area of sympatry¹². The hybrids are vigorous but have high, although not complete, male and female sterility^{13–15}. Experimental crosses indicate as many as 40% of hand-pollinated hybrid flowers will mature at least one seed and as many as 72% of crosses sired by hybrid pollen will set one seed¹⁵. Additionally, allozyme data show low levels of gene flow between these species of *Phlox*¹⁴. Although other traits may contribute to prezygotic isolation

(including possible reinforcement traits such as self-compatibility¹²), Levin demonstrated that the shift from light-blue to dark-red flowers in *P. drummondii* decreases interspecific hybridization by 66%, indicating that the change in flower colour substantially increases prezygotic reproductive isolation¹². Given the above estimates, the hybridization rate before the evolution of dark-red flower colour could have been as high as 28%, which, with low hybrid fitness, would presumably create strong selection to decrease hybridization.

We determined that the evolutionary transition from light-blue to dark-red flower colour in *P. drummondii* results from changes of large effect at two loci. F₂ populations derived from crosses between the allopatric colour variant (light blue) and the sympatric colour variant (dark red) segregate four discrete flower colours: dark blue, light blue, dark red and light red (Fig. 1a). Quantification of the spectral reflectance of 200 F₂ flowers, transformed into two-dimensional Commission internationale de l'éclairage (CIE) 1976 colour space¹⁶, followed by discriminant analysis verified our discrete classifications (Supplementary Fig. 1 and Supplementary Table 4a, b). The ratios of counts within these categories are very close to the 9:3:3:1 ratios expected from two loci with complete dominance ($\chi^2_{(3, N=618)} = 0.92$, $P = 0.8206$) (Supplementary Table 3). One locus, *H*, determines flower hue, with blue allele dominant to red, whereas the second locus, *I*, determines colour intensity, with the dark allele dominant to the light.

These two loci appear to determine the types and amounts of anthocyanin floral pigment produced in *P. drummondii*. Anthocyanin pigments, the final products of the well-characterized and highly conserved ABP, are derived from six common types of anthocyanidin by the addition of sugar and/or methyl moieties¹⁷. Less-hydroxylated anthocyanidins give rise to redder pigments whereas more-hydroxylated anthocyanidins give rise to bluer pigments¹⁷. Correspondingly, blue-flowered *P. drummondii* (*H*-) produce anthocyanins derived from both the less-hydroxylated cyanidin and peonidin pigments, as well as the more-hydroxylated malvidin pigment, whereas red flowers (*hh*) produce exclusively the less-hydroxylated pigments (Fig. 1b). The change in floral hue thus results from redirecting flux from the malvidin branch of the anthocyanin pathway to the cyanidin/peonidin branch (Fig. 1c). Individuals with increased colour intensity (*I*-) produce more pigment than *ii* individuals, without an effect on pigment composition (Fig. 1b and Supplementary Table 5a, b). These biochemical patterns, coupled with the structure of the ABP (Fig. 1c), suggest candidate genes for the hue and intensity loci.

The structure of the ABP (Fig. 1c) suggests that the loss of flux down the malvidin branch of the ABP might result from changes in one of three candidate genes: (1) loss of function or reduced expression of the gene coding for the branching enzyme flavanoid 3'-hydroxylase (F3'H); (2) alteration of the substrate specificity owing to a coding mutation in the gene for dihydroflavonol 4-reductase (DFR) making the enzyme unable to metabolize the malvidin precursor dihydromyricetin; and (3) a similar alteration of the substrate specificity of anthocyanidin synthase (ANS). To examine these possibilities, we analysed multiple F₂ populations for co-segregation of floral hue with single nucleotide polymorphism (SNP) markers in the candidate genes. In a total of 100

¹Department of Biology, Box 90338, Duke University, Durham, North Carolina 27708, USA.

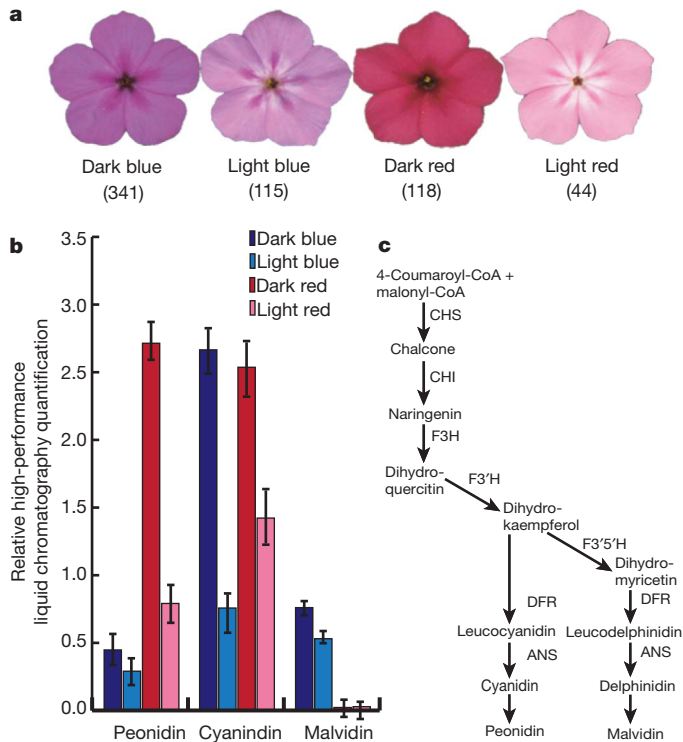


Figure 1 | Flower colour phenotypes in F_2 individuals. **a**, Representative pictures of the four flower colours in F_2 populations: dark blue, light blue, dark red, light red (from left to right). Total counts in F_2 populations indicated under each flower. Ratios of counts are similar to 9:3:3:1 ($\chi^2_{(3, N=618)} = 0.92$, $P = 0.8206$). **b**, Relative anthocyanidin pigment production results in variation in F_2 flower colour. Production of all three pigments is significantly different between hue classes. The light- and dark-red flowers produce no malvidin ($F_{(1,25)} = 134.03$, $P < 0.0001$), and significantly more peonidin ($F_{(1,25)} = 58.88$, $P < 0.0001$) and cyanidin ($F_{(1,25)} = 21.57$, $P = 0.0002$). The amount of pigment production is significantly different between classes of flower intensity, with high-intensity flowers producing more of each flower colour pigment ($F_{\text{malvidin}}(1,25) = 8.64$, $P = 0.0092$; $F_{\text{peonidin}}(1,25) = 32.11$, $P < 0.0001$; $F_{\text{cyanidin}}(1,25) = 66.74$, $P < 0.0001$). Standard errors are shown. See Supplementary Table 5a–c for full data and multivariate analysis of variance (MANOVA) results. **c**, A simplified schematic of the ABP, showing core enzymes to the right of the arrows, with the substrates and products indicated at the ends of the arrows. The two branches of the pathway active in *P. drummondii* produce two 3'-hydroxylated red pigments (cyanidin and peonidin) and one 5'-hydroxylated blue pigment (malvidin).

individuals there was perfect co-segregation between $F3'5'h$ and floral hue (Supplementary Table 7). Moreover, genotype at $F3'5'h$ explains 77% of the variation in flower hue.

This genetic association corresponds to a downregulation of $F3'5'h$ in red flowers. Among F_2 individuals, there is a nearly 100-fold decrease in $F3'5'h$ transcripts in red-flowered compared with blue-flowered individuals (Fig. 2a). In addition, red- and blue-flowered individuals collected from multiple populations throughout the range of *P. drummondii* have a comparable difference in $F3'5'h$ expression (Fig. 2b). These results demonstrate three features of the genetic change associated with the shift to red flowers: (1) variation in hue is associated with transcript level of $F3'5'h$; (2) genotype at $F3'5'h$ predicts transcript level; and (3) the expression difference evident between naturally occurring flower colour variants segregates as a single locus in F_2 individuals.

These patterns imply that expression variation at $F3'5'h$ is caused by variation in a *cis*-regulatory element. Allele-specific expression assays confirm this inference: in heterozygous (*Hh*) individuals, the red allele is almost completely downregulated (Fig. 2c). This allelic imbalance indicates a *cis*-regulatory change, whereas equal expression of the two alleles would indicate *trans*-acting regulation of expression¹⁸.

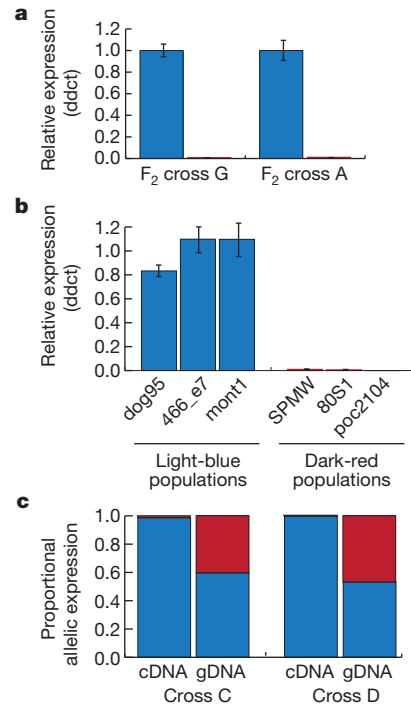


Figure 2 | Results of expression experiments on the hue locus ($F3'5'h$).

Quantitative PCR results showing relative expression of $F3'5'h$ in blue and red individuals in two F_2 populations (**a**) and six field populations (**b**). Transcripts of $F3'5'h$ in red floral tissue (shown in red) are significantly lower than levels detected in blue tissue (shown in blue) in both F_2 populations ($F_{(1,26)} = 250.89$, $P < 0.0001$) and field-collected individuals ($F_{(1,36)} = 30.21$, $P = 0.0053$). Delta-delta cycle-threshold (ddct) indicates relative transcript levels, and bars indicate two standard error units. **c**, Allele-specific expression of $F3'5'h$ in multiple heterozygous individuals (*Hh*) from two crosses. Each bar represents the relative contribution of the red allele (shown in red) and the blue allele (shown in blue) to the total expression detected in heterozygous individuals from F_2 families C and D. Relative allelic representation in complementary DNA (cDNA) from heterozygous individuals is significantly different from the relative allelic representation in genomic DNA (gDNA) ($F_{(1,32)} = 375.93$, $P < 0.0001$), indicating a *cis*-regulatory change controlling expression of the red allele. The genomic samples from heterozygous individuals show nearly equal allelic representation (0.5). See Supplementary Fig. 3 for experimental control data.

Based on the structure of the ABP, we identified three candidate genetic changes that could explain the increased pigment production in dark (*I*-) flowers: (1) a *cis*-regulatory change that increases production of an ABP enzyme with control over pathway flux; (2) increased enzymatic efficiency (through coding-sequence mutations) of a rate-controlling enzyme in the ABP; and (3) increased expression or function of an ABP transcription factor coordinately regulating the expression of several enzymes. To evaluate these possibilities, we cloned and sequenced genes coding for the core enzymes of the pathway (Fig. 1c), as well as an R2R3-Myb transcription factor orthologous to those known to regulate ABP enzymes in other angiosperms. Of these genes, the marker in R2R3-Myb co-segregates perfectly with flower colour intensity in 100 F_2 individuals (Supplementary Table 8) and genotype explains 71% of the intensity variation. These observations suggest *Myb* corresponds to the intensity locus.

This correspondence is further supported by quantification of *Myb* transcript in F_2 individuals. *Myb* expression is significantly higher in dark individuals than light individuals (Fig. 3a), demonstrating that both genotype and expression of this transcription factor are associated with variation in pigment production. Furthermore, field-collected individuals show the same association between expression of *Myb* and flower colour intensity (Fig. 3b). Finally, analysis of allele-specific expression in heterozygous (*Ii*) individuals show that the 'dark' allele is

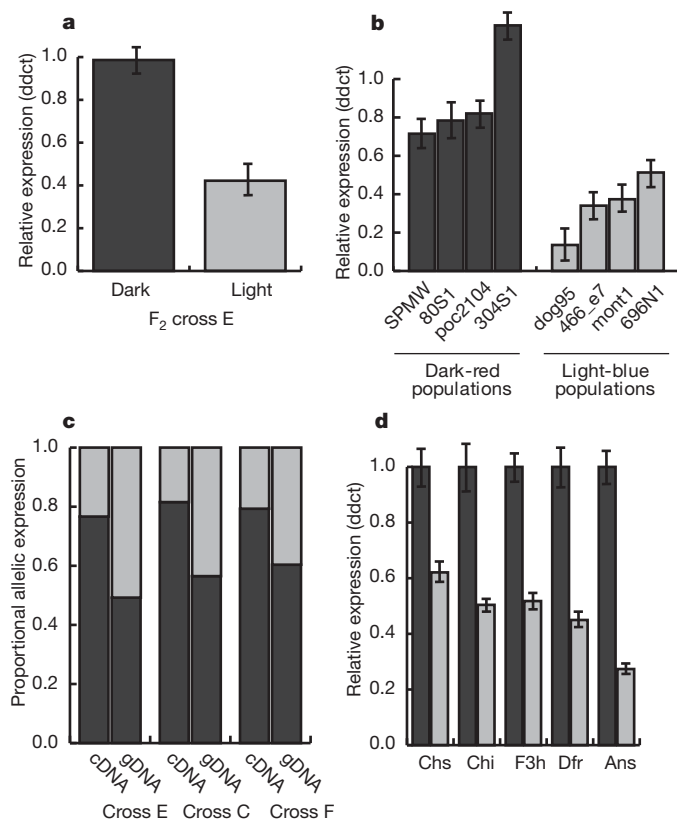


Figure 3 | Results of expression experiments of the intensity locus (*R2R3-Myb*). Relative expression of *R2R3-Myb* in dark and light F₂ individuals (**a**) and natural field populations (**b**). There is a significant upregulation associated with dark individuals relative to light both in F₂ individuals ($F_{(1,12)} = 8.11$, $P = 0.0173$) and in field-collected individuals ($F_{(1,50)} = 14.91$, $P = 0.0023$). **c**, Allele-specific expression indicates significantly different allelic representation in cDNA relative to gDNA in heterozygous individuals (*It*) from three F₂ families ($F_{(1,12)} = 116.74$, $P < 0.0001$). The over-representation of the dark allele (dark grey) relative to the light allele (light grey) indicates a *cis*-regulatory change. See Supplementary Fig. 4 for complete control data. **d**, There is significant upregulation of transcript levels of all core ABP enzymes in field-collected dark individuals (dark grey) relative to light individuals (light grey). $F_{chs(1,36)} = 5.19$, $P = 0.0305$; $F_{chi(1,36)} = 5.07$, $P = 0.032$; $F_{f3h(1,36)} = 15.81$, $P = 0.0004$; $F_{dfr(1,36)} = 13.45$, $P = 0.001$; $F_{ans(1,36)} = 36.7$, $P < 0.0001$. For full MANOVA results see Supplementary Table 10. Standard errors are indicated on graphs.

upregulated relative to the 'light' allele (Fig. 3c), a pattern indicative of a *cis*-regulatory change at the *Myb* locus.

In all species that have been examined, this *Myb* coordinately activates several, if not all, of the ABP enzyme-coding genes¹⁹. Thus we expect that if changes in expression of *Myb* influence pigment intensity, there should be correlated expression changes in the ABP enzyme-coding genes. This expectation was realized: all five core-enzyme genes exhibited significant upregulation in dark flowers relative to light flowers (Fig. 3d). These results indicate that a *cis*-regulatory mutation at the *R2R3-MYB* transcription factor causes the increased colour intensity of *P. drummondii* flowers in sympatric populations.

Our investigation has provided the first identification of genetic changes causing reinforcement of species boundaries. Ideally, we would like to have confirmed the identity of the hue and intensity loci by either fine mapping or transformation, but this is not currently feasible in non-model systems like *P. drummondii*. Nevertheless, we are confident that we have identified the correct genes. Not only do *F3'5'h* and *Myb* co-segregate perfectly with the hue and intensity loci, respectively, but both exhibit changes in expression levels in the directions that are expected to produce the observed phenotypic changes. Moreover, the alternative possibility that the changes are due to linked

transcription factors is ruled out by differences in allele-specific expression.

We have shown that reinforcement may involve changes in a few genes, each change having a large phenotypic effect. Our results expand upon two previous analyses of the genetic architecture of reinforcement, which report the involvement of a small number of quantitative trait loci^{10,11}. This simple genetic architecture is consistent with theoretical expectations, which indicate that selection for reinforcement is most likely to result in increased reproductive isolation when the phenotypic effect of the assortative mating allele is large and selection for reproductive isolation is strong^{7,20}.

Ortiz-Barrientos *et al.*¹⁰ suggest that reinforcing reproductive isolation should be inherited as a dominant trait because of Haldane's sieve (the greater probability of a new dominant adaptive mutation reaching fixation than a recessive mutation²¹). The intensity locus fits this expectation, with the derived dark allele dominant to the light allele. In contrast, the hue locus shows the reverse pattern, with the derived red allele recessive to the ancestral blue allele. Although it is easier for selection to increase the frequency of a novel dominant allele, Haldane's sieve can be overcome with strong selection²¹. Additionally, probability of fixation has been found to be independent of dominance when adaptations are not new mutations but are standing genetic variation²². Either of these possibilities could explain the fixation of the red allele in sympatry.

Much work investigating the process of speciation focuses on categorizing traits as a particular type of reproductive isolation mechanism (that is, prezygotic or postzygotic) (reviewed by Lowry *et al.*²³ and Nosil *et al.*²⁴). Recently it has become clear that a single trait can affect multiple types of reproductive isolation^{25,26}, but it remains unclear how commonly this occurs. Although the most obvious effect of floral-colour evolution in *P. drummondii* is increased premating isolation with *P. cuspidata*, the associated genetic changes may also have led to increased postzygotic isolation. Hybrids between the dark-red-flowered *P. drummondii* and the light-blue-flowered *P. cuspidata* have dark-blue flowers, which differ from the two parental species. Levin has shown that pollinators discriminate against this intermediate hybrid flower colour²⁷. Although further experimental tests are required, this discrimination probably reduces male outcross success and possibly seed set, causing reduced fitness of the hybrids beyond that associated with intrinsic postzygotic isolation. Our work shows how the pattern of genetic dominance can influence whether a trait can contribute to multiple types of isolation. This hypothesized extrinsic postzygotic isolation arises only because the novel allele at the intensity locus is dominant, whereas the novel allele at the hue locus is recessive. Had the novel alleles at both loci exhibited the same dominance, hybrids would have had the same flower colour as one of the parents.

Finally, both of the genetic changes contributing to reinforcement in *P. drummondii* involve *cis*-regulatory mutations, rather than functional (coding-sequence) mutations in pathway enzymes. The types of mutation described here are similar to those described for other cases of flower colour evolution in which reinforcement is not known to be involved^{28–30}. These similarities suggest that the genetic basis of reinforcement may be comparable to that of adaptations not associated with reinforcement. However, only by indentifying the genetic basis of reinforcement in other systems can we understand what aspects of the molecular genetic basis of these traits are generally important for reinforcement and which are unique to a particular system.

METHODS SUMMARY

F₂ populations were created from crossing parents collected from allopatry with those collected from sympatry (Supplementary Table 1). F₂ individuals were phenotyped for flower colour and sampled for DNA and RNA extractions. Anthocyanidins were extracted from floral tissue using standard protocol and identified and quantified using high-performance liquid chromatography³⁰. Candidate genes were amplified from parental individuals to identify SNPs, and F₂ individuals were genotyped. The Roche Universal ProbeLibrary system was used to quantify expression of all candidate

genes from F_2 and field-collected populations. Pyrosequencing technology was used to quantify allele-specific expression in individuals heterozygous at flower colour loci¹⁸.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 August; accepted 2 November 2010.

Published online 9 January 2011.

- Noor, M. A. F. & Feder, J. L. Speciation genetics: evolving approaches. *Nature Rev. Genet.* **7**, 851–861 (2006).
- Wolf, J. B. W., Lindell, J. & Backstrom, N. Speciation genetics: current status and evolving approaches. *Phil. Trans. R. Soc. B* **365**, 1717–1733 (2010).
- Dobzhansky, T. *Genetics and the Origin of Species* (Columbia University Press, 1937).
- Howard, D. J. in *Hybrid Zones and the Evolutionary Process* (ed. Harrison, R. G.) 46–69 (Oxford University Press, 1993).
- Servedio, M. R. & Noor, M. A. F. The role of reinforcement in speciation: theory and data. *Annu. Rev. Ecol. Syst.* **34**, 339–364 (2003).
- Butlin, R. Speciation by reinforcement. *Trends Ecol. Evol.* **2**, 8–13 (1987).
- Felsenstein, J. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. *Evolution* **35**, 124–138 (1981).
- Ortiz-Barrientos, D., Greal, A. & Nosil, P. The genetics and ecology of reinforcement implications for the evolution of prezygotic isolation in sympatry and beyond. *Year Evol. Biol.* **2009**, 156–182 (2009).
- Pfennig, K. S. & Pfennig, D. W. Character displacement: ecological and reproductive responses to a common evolutionary problem. *Q. Rev. Biol.* **84**, 253–276 (2009).
- Ortiz-Barrientos, D., Counterman, B. A. & Noor, M. A. F. The genetics of speciation by reinforcement. *PLoS Biol.* **2**, 2256–2263 (2004).
- Saether, S. A. *et al.* Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* **318**, 95–97 (2007).
- Levin, D. A. Reproductive character displacement in *Phlox*. *Evolution* **39**, 1275–1281 (1985).
- Levin, D. A. Hybridization between annual species of *Phlox* – population structure. *Am. J. Bot.* **54**, 1122–1128 (1967).
- Levin, D. A. Interspecific hybridization, heterozygosity and gene exchange in *Phlox*. *Evolution* **29**, 37–51 (1975).
- Ruane, L. G. & Donohue, K. Pollen competition and environmental effects on hybridization dynamics between *Phlox drummondii* and *Phlox cuspidata*. *Evol. Ecol.* **22**, 229–241 (2008).
- Gonnet, J. F. CIELab measurement, a precise communication in flower color: an example with carnation (*Dianthus Caryophyllus*) cultivars. *J. Hort. Sci.* **68**, 499–510 (1993).
- Holton, T. A. & Cornish, E. C. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* **7**, 1071–1083 (1995).
- Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85–88 (2004).
- Koes, R., Verweij, W. & Quattrocchio, F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* **10**, 236–242 (2005).
- Kirkpatrick, M. Reinforcement and divergence under assortative mating. *Proc. R. Soc. Lond. B* **267**, 1649–1655 (2000).
- Haldane, J. B. S. A mathematical theory of natural and artificial selection, part 1. *Trans. Camb. Phil. Soc.* **23**, 19–41 (1924).
- Orr, H. A. & Betancourt, A. J. Haldane's sieve and adaptation from the standing genetic variation. *Genetics* **157**, 875–884 (2001).
- Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A. & Willis, J. H. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Phil. Trans. R. Soc. B* **363**, 3009–3021 (2008).
- Nosil, P., Vines, T. H. & Funk, D. J. Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**, 705–719 (2005).
- Dambroski, H. R. *et al.* The genetic basis for fruit odor discrimination in *Rhagoletis* flies and its significance for sympatric host shifts. *Evolution* **59**, 1953–1964 (2005).
- Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, e1000500 (2010).
- Levin, D. A. The exploitation of pollinators by species and hybrids of *Phlox*. *Evolution* **24**, 367–377 (1970).
- Hoballah, M. E. *et al.* Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell* **19**, 779–790 (2007).
- Schwinn, K. *et al.* A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* **18**, 831–851 (2006).
- Des Marais, D. L. & Rauscher, M. D. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution* **64**, 2044–2054 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Des Marais, J. Tung, S. Johnsen and T. Juenger for technical advice, D. Levin for assistance in locating natural populations, and M. Noor for comments on the manuscript. This work was supported by a National Science Foundation grant to M.D.R. and a National Science Foundation Doctoral Dissertation Research Improvement Grant to R.H. R.H. was supported in part by the National Science Foundation Graduate Research Fellowship Program.

Author Contributions R.H. and M.D.R. designed the project; R.H. performed the experiments and the analyses; R.H. and M.D.R. wrote the paper.

Author Information The DNA sequences reported here are deposited in GenBank under accession numbers HQ127319–HQ127344 and HQ323688–HQ323691. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.H. (robin.hopkins@duke.edu) or M.D.R. (mrausher@duke.edu).

METHODS

F₂ crosses. Seeds were collected from plants in six populations (Supplementary Table 1) in the spring of 2006, and germinated and grown at the Duke University greenhouses. Seeds were soaked for 48 h in 500 p.p.m. gibberellic acid, planted in Metro-Mix 360 (Sun Gro Horticulture), and stratified for 6 days at 4 °C. Plants were transplanted at the four-true-leaf stage into a 50:50 mix of Fafard 4P (Conrad Fafard) and PermaTill One Time (Carolina Stalite Company). Light-blue plants were crossed with dark-red plants to form F₁ seeds. F₁ seeds were grown as above and self-fertilized to create F₂ populations. Self-incompatibility was overcome using bud pollination.

Flower colour phenotyping. All F₂ individuals were categorically phenotyped for flower colour, and a χ^2 test was used to determine if the ratio of categorical flower colour counts differed from 9:3:3:1 (Supplementary Table 3). A subset of individuals were quantitatively phenotyped with a StellarNet EPP2000C spectrometer with a SL1 visible light source and a SL5 deuterium + halogen ultraviolet-visible light source (StellarNet). Raw reflectance was transformed into two-dimensional CIE 1976 LUV colour space¹⁶ (Supplementary Fig. 1). This transformation creates two axes of variation: the *x* axis corresponds to *u'* and the *y* axis corresponds to *v'*. The white point is located at (*v'* = 0.4683305, *u'* = 0.197833). The *v'* coordinate represents a measure of hue, and distance from the white point is a quantitative measure of intensity. We used canonical discriminant analysis to confirm that flower colour grouped into four discrete categories (Supplementary Table 4a, b). Anthocyanidins were extracted from a total of 26 individuals (12 from cross A, 14 from cross B) using standard methods of dissolving petal tissue in HCl followed by isoamyl alcohol extractions³⁰. Pigments were identified using high-performance liquid chromatography as described in Des Marais and Rauscher³⁰. Pigments were quantified by calculating the area under the high-performance liquid chromatography peak and scaled to the corresponding standards. We used a MANOVA to determine if pigment amount differed by intensity, hue, family and all interactions of main effects. We used subsequent individual ANOVA models to determine significance of each of the above effects on amount of individual pigments (Supplementary Table 5a, b). All analyses used SAS software version 9.1 (SAS Institute).

Genetic association. Leaf tissue was collected from F₂ and parental individuals. A modified cetyltrimethylammonium bromide (CTAB) extraction was used to isolate the DNA (as in Kelly and Willis³¹ but with an additional 2% Triton X-100 added to the CTAB solution and a 3 M sodium acetate wash after ethanol precipitation). Genes in the ABP were amplified, first using degenerate primers designed from orthologous genes in closely related species and then with species-specific primers in subsequent amplifications (Supplementary Table 6). Parental sequences for each gene were used to identify SNPs segregating in the F₂ populations. A subset of F₂ individuals were genotyped at each ABP gene. For those genes that showed an association, subsequent F₂ individuals were genotyped. χ^2 tests were used to confirm associations between genotype and phenotype (Supplementary Tables 7 and 8). A mixed-model ANOVA with F₂ family as a random effect was used to determine how much quantitative flower colour variation was explained by *F3'5'h* and the *R2R3-Myb* genotype. For this analysis we used flower colour reflectance transformed into CIE colour space to determine quantitative measures of hue and intensity (see above). The *y* axis, corresponding to the value of *v'*, is the measure of hue and the distance each flower colour point is from the white point (*v'* = 0.4683305, *u'* = 0.197833) is the measure of intensity. All analyses used SAS software version 9.1 (SAS Institute).

RNA expression analyses. All expression analyses were performed on flower-bud tissue collected approximately 2 days before opening. RNA was extracted from individuals using the SpectrumTM Plant Total RNA Kit (Sigma-Aldrich). The Roche Applied Science Universal ProbeLibrary (Roche Diagnostics) was used to quantify expression of each gene in the ABP³². Probes sites and primers were designed using the online design centre (<http://www.roche-applied-science.com/sis/rtpcr/upl/ezhome.html>) (Supplementary Table 9). A Thermo Scientific Verso 1-Step RT-qPCR kit was used to amplify according to the manufacturer's instructions (Fisher Scientific). Twenty-six F₂ individuals were used in the *F3'5'h* expression assay. Twelve F₂ individuals were used in the *Myb* expression assay. Two replicate reactions were performed for each sample and the average cycle-threshold value was used in all analyses. Efl- α was amplified in each sample to control for total amount of RNA in each extraction. Raw expression data were analysed as in Rieu and Powers³³. We used a mixed-model ANOVA to detect a significant difference in expression in candidate genes between colour groups. F₂ family was used as a random effect in the model.

We collected seeds from natural populations of *P. drummondii* in both allopatry and sympatry (Supplementary Table 2), grew them in the Duke University greenhouse and extracted RNA from bud tissue as described above. Thirty-six individuals were used from field-collected population expression assays, with an additional 14 individuals from two additional populations for the *Myb* assay. Transcript levels of all genes in the ABP were quantified as above. We used a MANOVA to determine if flower colour intensity has an effect on expression of non-causal ABP genes (*Chs*, *Chi*, *F3h*, *Dfr*, *Ans*). Subsequent ANOVAs were used to determine if each individual gene shows a significant effect of intensity on expression (Supplementary Table 10). All analyses were performed using SAS software version 9.1 (SAS Institute).

Allele-specific expression. RNAs from F₁ individuals and heterozygous F₂ individuals were used to quantify allele-specific expression at both the *F3'5'h* and the *Myb* genes. A schematic of the assay design (Supplementary Fig. 2) shows the SNP identity, as well as amplification and sequencing primer sequences. RNA was extracted as described above from 28 individuals for the *Myb* assays (11 heterozygotes and seven homozygous controls) and from eight individuals for the *F3'5'h* assay (six heterozygotes and two homozygous controls). A reverse transcription reaction was performed using InvitrogenTM SuperScript II (Life Technologies) according to the manufacturer's instructions. DNA was extracted from leaf tissue as described above. For each individual, two replicate DNA and four replicate cDNA PCRs were run. No-template controls and no-sequencing-primer controls were performed as well. Pyrosequencing reactions were run on all samples using PyroMARKTM Q96ID (Qiagen)^{18,34}. Experiments on both genes were independently replicated.

31. Kelly, A. J. & Willis, J. H. Polymorphic microsatellite loci in *Mimulus guttatus* and related species. *Mol. Ecol.* **7**, 769–774 (1998).
32. Mouritzen, P. et al. The ProbeLibrary™-expression profiling 99% of all human genes using only 90 dual-labeled real-time PCR probes. *Biotechniques* **37**, 492–495 (2004).
33. Rieu, I. & Powers, S. J. Real-time quantitative RT-PCR: design, calculations, and statistics. *Plant Cell* **21**, 1031–1033 (2009).
34. Ahmadian, A. et al. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280**, 103–110 (2000).

Reversing pathological neural activity using targeted plasticity

Navzer D. Engineer^{1,2}, Jonathan R. Riley¹, Jonathan D. Seale¹, Will A. Vrana¹, Jai A. Shetake¹, Sindhu P. Sudhanagunta¹, Michael S. Borland¹ & Michael P. Kilgard¹

Brain changes in response to nerve damage or cochlear trauma can generate pathological neural activity that is believed to be responsible for many types of chronic pain and tinnitus^{1–3}. Several studies have reported that the severity of chronic pain and tinnitus is correlated with the degree of map reorganization in somatosensory and auditory cortex, respectively^{1,4}. Direct electrical or transcranial magnetic stimulation of sensory cortex can temporarily disrupt these phantom sensations⁵. However, there is as yet no direct evidence for a causal role of plasticity in the generation of pain or tinnitus. Here we report evidence that reversing the brain changes responsible can eliminate the perceptual impairment in an animal model of noise-induced tinnitus. Exposure to intense noise degrades the frequency tuning of auditory cortex neurons and increases cortical synchronization. Repeatedly pairing tones with brief pulses of vagus nerve stimulation completely eliminated the physiological and behavioural correlates of tinnitus in noise-exposed rats. These improvements persisted for weeks after the end of therapy. This method for restoring neural activity to normal may be applicable to a variety of neurological disorders.

Damage to the peripheral nervous system causes plasticity in multiple regions of the central nervous system. Significant changes have been reported in map organization, spontaneous activity, neural synchronization and stimulus selectivity². The ideal method of testing whether map plasticity or some other form of plasticity is directly responsible for chronic pain and tinnitus would be to reverse the plasticity and evaluate the perceptual consequence.

Recent attempts to use sensory exposure or discrimination training to reverse the map changes in individuals with tinnitus or chronic pain have provided some temporary relief^{6,7}. Although the clinical benefits were limited, these studies provide some support for the hypothesis that neural plasticity could be used to treat these conditions. It is possible that a long-lasting reversal of the pathological plasticity in these patients would provide significant relief.

Studies in animals have shown that repeatedly pairing sensory stimuli with electrical stimulation of the cholinergic nucleus basalis generates powerful and long-lasting changes in cortical organization⁸. In principle, this method could be used to reverse the effect of pathological plastic changes that are associated with tinnitus and chronic pain^{1–3,6}. However, nucleus basalis stimulation is highly invasive and, thus, not practical for clinical use. We have developed a less invasive method for generating targeted neural plasticity by pairing vagus nerve stimulation (VNS) with sensory inputs, and have demonstrated a potential clinical application.

VNS triggers the release of neuromodulators known to promote plastic changes. The efficacy of VNS in enhancing plasticity seems to lie in the synergistic action of multiple neuromodulators acting in the cerebral cortex and other brain regions⁹. VNS improves learning and memory of associated events in rats and humans using identical VNS parameters¹⁰.

Our study tests the hypothesis that the pairing of VNS with tones could be used to drive neural plasticity that would reverse the behavioural correlate of tinnitus in noise-exposed rats. The first set of experiments confirms that repeatedly pairing a single tone frequency with VNS is sufficient to generate specific and long-lasting changes in cortical maps.

The rationale for our tinnitus therapy is that increasing the number of cortical neurons tuned to frequencies other than the tinnitus frequency ought to reduce the overrepresented tinnitus frequency. The second set of experiments confirms that repeatedly pairing a range of tone frequencies with VNS can be used to reverse the behavioural and neural correlates of tinnitus in noise-exposed rats.

In our first set of experiments, we sought to evaluate whether pairing VNS with tones can generate precise, long-lasting and large-scale changes in the frequency representation in the cortex, as we found for nucleus basalis stimulation. We paired VNS with a 9-kHz, 60-dB SPL tone ($n = 8$ rats) or a 19-kHz, 50-dB SPL tone ($n = 5$ rats) for 20 days (SPL, sound pressure level), 300 times per day in normal-hearing rats with cuff electrodes implanted on the left cervical vagus nerve (Methods). The VNS–tone pairing procedure was identical to earlier tone pairing procedures with nucleus basalis, ventral tegmentum or locus coeruleus stimulation that generate long-lasting map plasticity^{8,11,12}. VNS parameters (30 Hz, 0.8 mA) were similar to the parameters used in previous rat and human VNS studies, except that the duration of stimulation and the widths of individual pulses were reduced by 60-fold and fivefold, respectively (Methods and Supplementary Fig. 2). The 0.5 s of VNS used in this study was sufficient to reduce the amplitude of the cortical electroencephalogram briefly (Supplementary Fig. 3 and supplemental data). Twenty-four hours after the last VNS–tone pairing session, we used standard micro-electrode mapping techniques to document frequency map plasticity. VNS–tone pairing caused a 70–79% increase in the number of primary auditory cortex (A1) sites with a characteristic frequency near the paired tone frequency (Fig. 1). This result confirms our hypothesis that VNS–tone pairing can be used to direct map plasticity lasting more than 24 h.

Pairing VNS with sensory stimuli is a potentially attractive method of modifying neural circuits without significant side effects. VNS is well tolerated in the 50,000 patients who currently receive VNS therapy for epilepsy or depression¹³. By pairing tones with brief trains of VNS, we have been able to alter cortical frequency maps significantly in rats using only 1% of the VNS that is delivered clinically (that is, 30 s every 5 min, 24 h per day) for epilepsy treatment in humans.

Having demonstrated that VNS can be used to generate specific and long-lasting map plasticity, in our second set of studies we sought to evaluate whether VNS-directed plasticity could be adapted to renormalize pathological plasticity and eliminate tinnitus. Exposure to intense, high-frequency noise is known to generate an overrepresentation of mid-frequency tones, degrade frequency selectivity and increase excitability and synchronization of auditory neurons^{14–16}. We induced noise trauma by exposing rats to 1 h of 115-dB SPL, octave-band noise centred at 16 kHz (ref. 17; Methods). Auditory brainstem responses were used to confirm the effects of the noise exposure on hearing threshold, including temporary deafness for frequencies above 8 kHz and a long-lasting increase of auditory brainstem response thresholds and latency¹⁸ (Supplementary Figs 4 and 5). After noise exposure, twice as many A1 recording sites were tuned to frequencies between 2 and 4 kHz in comparison with naive controls ($35 \pm 7\%$ versus $14 \pm 2\%$, $P < 0.05$), and very few neurons responded to frequencies above

¹Cortical Plasticity Laboratory, Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, Texas 75080, USA. ²MicroTransponder Inc., 2802 Flintrock Trace, Suite 225, Austin, Texas 78738, USA.

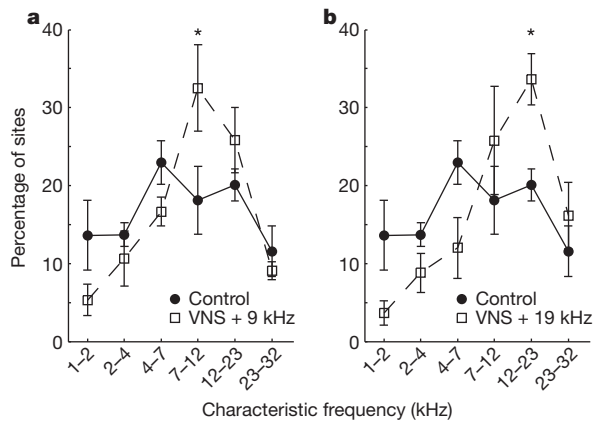


Figure 1 | VNS-tone pairing causes map plasticity. Repeatedly pairing VNS with a tone increases the number of A1 recordings sites tuned to the paired frequency. **a**, VNS was paired with a 9-kHz tone 6,000 times over 20 days in eight rats. **b**, VNS was paired with a 19-kHz tone in five rats. This group heard 4-kHz tones equally often but without VNS pairing. Asterisks indicate significant ($P < 0.05$) increases in the fraction of A1 sites with characteristic frequencies near the paired tone. Error bars, s.e.m. This result in normal-hearing rats suggested that VNS-tone pairing might be used to reverse the map distortions induced by exposure to intense noise.

23 kHz ($1.7 \pm 1\%$ versus $11.5 \pm 3\%$, $P < 0.01$). The average frequency bandwidth of A1 neurons increased by 21% (1.75 ± 0.04 versus 1.47 ± 0.03 octaves at 10 dB above threshold, $P < 0.00001$), and the average number of spikes evoked by a tone within each site's receptive field increased by 30% (4.3 ± 0.1 versus 3.3 ± 0.1 , $P < 0.00001$). The average spontaneous rate increased by 23% (17.7 ± 0.6 versus 14.3 ± 0.4 Hz, $P < 0.00001$). The degree of synchronization during silence measured using the correlation coefficient between multiunit activity recorded at nearby sites was significantly increased (1.7 ± 0.01 versus 0.19 ± 0.01 synchronous spikes per second of silence, $P < 0.05$; Methods). These changes in frequency tuning and synchronization are similar to the physiological changes observed after noise exposure that have been proposed to be directly responsible for tinnitus^{2,19}. Earlier studies using several different methods have documented that noise exposure can generate behavioural correlates of tinnitus near the low-frequency edge of the noise trauma^{17,20–22}. However, few studies have directly compared neurophysiology and behavioural observations from the same animals^{20,23}. It was therefore of great interest to us to relate noise-induced plasticity to perceptual disturbances.

Each of the eighteen noise-exposed rats used in this study was significantly impaired in its ability to detect a gap in narrowband noise centred on 8 or 10 kHz, but showed no impairment when the gap occurred in narrowband noise centred on 2 or 4 kHz or in broadband noise (Fig. 2, 4 weeks after exposure). Several studies have concluded that a frequency-specific impairment in gap detection is a likely sign that noise-exposed rats experience a mid-frequency tinnitus percept which fills the silent gaps^{17,23} (Methods and Supplementary Figs 6–9). Although it is not possible to evaluate the subjective experience of rats definitively, the gap impairment has been taken as a possible behavioural correlate of tinnitus.

Map distortion and tuning curve broadening (but not changes in spontaneous activity or synchronization) were significantly correlated with the degree of gap impairment in untreated noise-exposed rats ($R > 0.7$ (Pearson correlation coefficient), $P < 0.05$, $n = 8$ sham rats; Figs 3a, b and 4a–d and Supplementary Fig. 13). These correlations must be interpreted with caution because any variability in the initial cochlear trauma could generate a correlation between neural and behavioural changes even in the absence of a causal relationship. Though still not definitive, the best test for a causal relationship would be to reverse specifically the plasticity generated by noise exposure and document the reversal of the gap detection impairment.

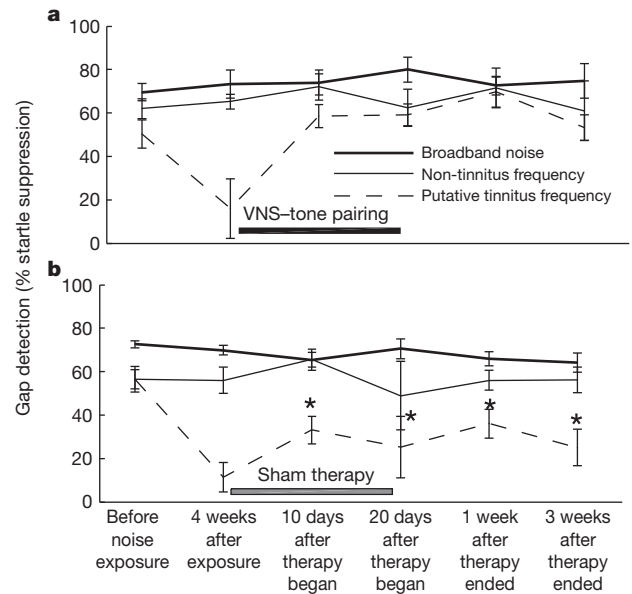


Figure 2 | VNS/multiple tone pairing eliminates the behavioural correlate of tinnitus. Four weeks after noise exposure, each of the rats in both groups was unable to detect a gap in one or more of the narrowband noises tested ($P > 0.05$; Supplementary Fig. 8b). The frequency with the greatest impairment four weeks after noise exposure is the putative tinnitus frequency for each rat. For both groups, gap detection at the putative tinnitus frequency was significantly impaired in comparison to broadband noise ($P < 0.05$). The gap detection at the non-tinnitus frequency is based on gap detection in 16-kHz narrowband noise. **a**, Gap detection at the putative tinnitus frequency (dotted line) improved significantly after ten days of VNS-tone pairing, and the improvement persisted at least until the acute physiology experiment ($n = 5$ rats). **b**, The sham group ($n = 9$ rats) continued to be impaired. Two sham rats did not contribute data at the non-tinnitus frequency because they showed gap impairments at 16 kHz (as well as 8 and 10 kHz) four weeks after noise exposure. Black and grey horizontal bars represent duration of VNS and sham therapy, respectively. Asterisks represent significant differences ($P < 0.05$) in gap detection at the putative tinnitus frequency between VNS therapy and sham therapy rats. Error bars, s.e.m.

We speculated that pairing VNS with randomly interleaved pure tones that span the rat hearing range, but exclude the overrepresented frequencies, could decrease the cortical representation of the excluded frequencies²⁴. We also expected that pairing multiple tone frequencies with VNS ('VNS/multiple tone' pairing) would increase frequency selectivity and decrease synchronization as in our earlier nucleus basalis stimulation experiments²⁵. We quantified behavioural and physiological correlates of tinnitus in noise-exposed rats and then tested whether pairing VNS with multiple tone frequencies could reverse the pathological plasticity and eliminate the perceptual disturbance in these rats.

VNS was repeatedly paired with multiple pure tones 300 times per day for 18 days in seven noise-exposed rats with impaired gap detection for mid-frequency sounds (Methods). Because we found that gap impairment occurred at 8–10 kHz, we selected the frequency of each randomly interleaved tone to be 1.3, 2.2, 3.7, 17.8 or 29.9 kHz. This pairing procedure was chosen because previous studies suggest it would reduce the cortical response to mid-frequency tones, increase frequency selectivity and decrease cortical synchronization^{2,25}. After ten days of therapy, each of the seven rats showed a significant startle reduction in cued trials relative to uncued trials for every frequency tested ($P < 0.05$; Fig. 2a and Supplementary Fig. 9a). Thus, pairing of VNS with multiple tones reversed the behavioural effect of noise exposure, which suggests that the rats' presumed tinnitus was no longer present. In contrast, rats in the sham therapy group showed a consistent

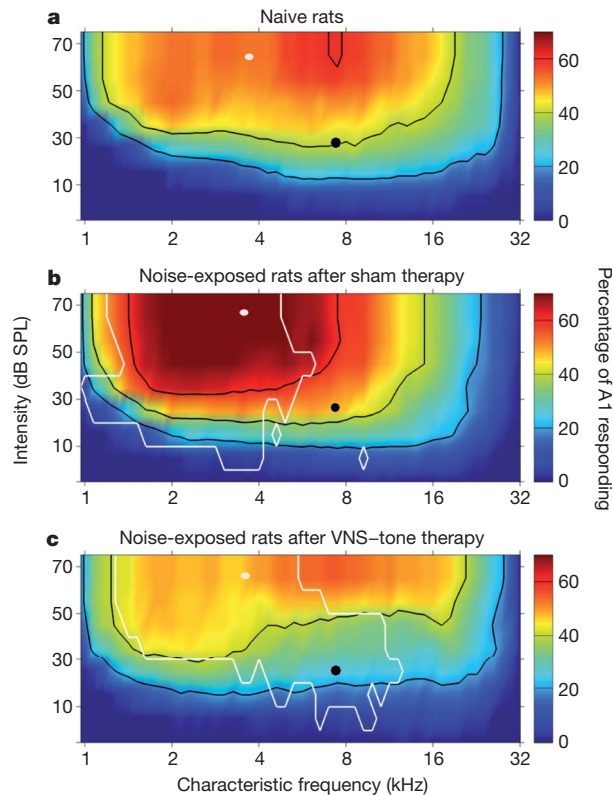


Figure 3 | VNS/multiple tone pairing reverses map distortion. The increased response of A1 neurons to tones following noise exposure is reversed by VNS/multiple tone pairing. **a**, Colour indicates the percentage of A1 neurons in naive rats that respond to a tone of any frequency and intensity combination. **b**, Percentage of A1 neurons that respond to each tone in noise-exposed rats that received sham therapy. **c**, Percentage of A1 neurons that respond to each tone in noise-exposed rats that received the VNS/multiple tone therapy. Black contour lines indicate 20, 40, and 60% responses. The white lines in **b** surround the regions of tones that are significantly increased ($P < 0.01$) in comparison with naive rats. The white lines in **c** indicate significant decreases ($P < 0.01$) in comparison with noise-exposed sham therapy rats. The filled white circles indicate the tone for which the increase in the number of cortical neurons was greatest, which is used to quantify the degree of map distortion in Fig. 4a, b. The filled black circles indicate the tone for which the proportional increase was greatest.

impairment in their ability to detect gaps in the putative tinnitus frequency (Fig. 2b). Each of the nine rats that received sham therapy (tones with no VNS, VNS with no tones or no therapy; $n = 4, 2, 3$ rats, respectively) did not show a significant startle reduction in cued trials ($P > 0.05$; Supplementary Fig. 9b) for at least one of the frequencies tested at each time point.

In the rats that received VNS paired with multiple tones, the impairment in gap detection was also eliminated when measured one day, one week and three weeks after the end of the therapy. This impairment was maintained in all three control groups at every time point tested (Supplementary Figs 9 and 10). These results indicate that pairing VNS with multiple tone frequencies is sufficient to eliminate the gap impairment induced by noise exposure (Supplementary Fig. 11). This is the first method reported to generate a long-lasting reversal of a behavioural correlate of chronic tinnitus.

Three weeks after the end of VNS/multiple tone pairing or sham therapy, we evaluated the physiological properties of the auditory cortex of each rat to determine whether the restored behaviour in the treated group was due to renormalization of the auditory cortex. After VNS/multiple tone pairing, most of the A1 properties that were degraded by noise exposure returned to normal levels. For example, the proportion of A1 neurons with characteristic frequencies between 12 and 23 kHz was indistinguishable from that in naive controls after VNS/multiple tone treatment (naive, $20 \pm 2\%$; sham, $15 \pm 5\%$; therapy, $30 \pm 9\%$; Supplementary

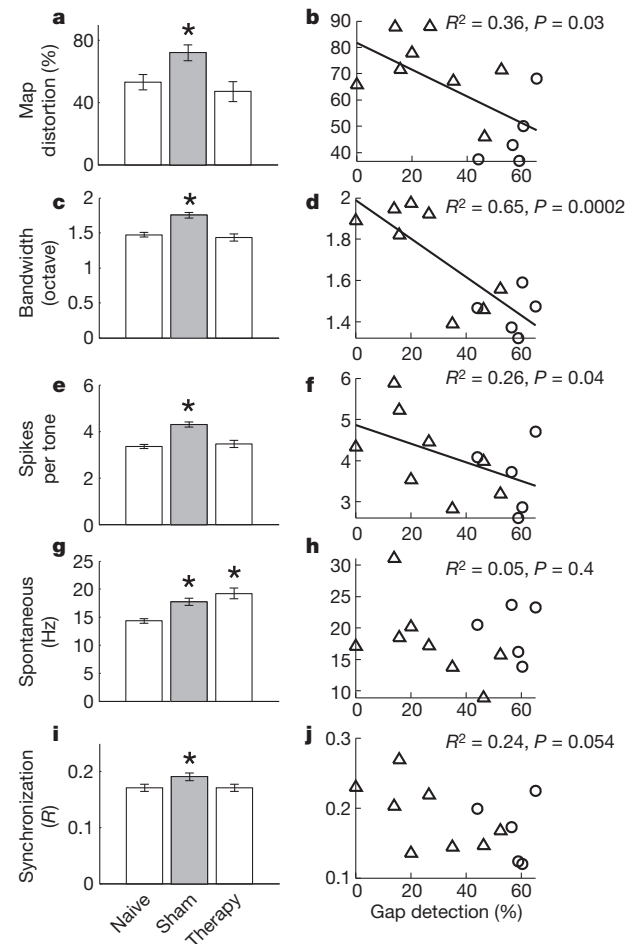


Figure 4 | Neurophysiological properties of naive, sham and therapy rats.

a, c, e, g, i, Noise exposure caused a significant map distortion (**a**), decreased frequency selectivity (**c**), increased the tone-evoked response (**e**), increased the spontaneous rate (**g**) and increased the degree of cortical synchronization (**i**). VNS/multiple tone pairing returned each of these parameters, except spontaneous activity, to normal levels. **b, d, f**, Map organization (**b**), frequency selectivity (**d**) and tone-evoked response strength (**f**) were all correlated with the degree of gap impairment in individual rats. **h, j**, Spontaneous activity (**h**) and synchronization (**j**) were not significantly correlated with gap impairment. Each rat's gap detection ability was quantified as the average gap detection at the putative tinnitus frequency of each rat, averaged across the four time points collected after the beginning of therapy (Fig. 2). Error bars, s.e.m. Asterisks represent significant differences compared with naive rats (P values as indicated). Triangles and circles represent rats from the sham and therapy groups, respectively.

Figs 12 and 13a). The proportion of A1 neurons responding to 4-kHz, 70-dB SPL tones significantly increased relative to naive controls in sham rats and returned to normal levels in rats that had received the therapy three weeks earlier (naive, $45.4 \pm 5.0\%$; sham, $74.1 \pm 7.6\%$; therapy, $49.1 \pm 6.6\%$; Figs 3 (white circles) and 4a). The degree of low-frequency map distortion was positively correlated with the degree of gap impairment observed in individual rats (Fig. 4b and Supplementary Fig. 13b). The percentage of cortex responding to 8-kHz, 30-dB SPL tones (Fig. 3, black circles) was also well correlated with the gap detection impairment ($R^2 = 0.51$, $P = 0.006$). These results support the earlier hypothesis that changes in cortical maps are causally related to tinnitus^{4,26}.

VNS/multiple tone pairing reversed the increase in the width of frequency tuning of A1 multiunit activity (that is, decreased frequency selectivity) observed in noise-exposed rats (Fig. 4c). The bandwidth (measured at 10, 20, 30 or 40 dB above threshold) averaged across all A1 sites was highly correlated with the degree of gap impairment (Fig. 4d and Supplementary Fig. 14), thus supporting the earlier hypothesis that decreased frequency selectivity is causally related to tinnitus²⁷.

VNS/multiple tone pairing reversed the increase in cortical excitability observed in noise-exposed rats (Fig. 4e). The average number of spikes evoked by tones within each site's receptive field was weakly correlated with the degree of impairment of gap detection (Fig. 4f), supporting the earlier hypothesis that tinnitus is related to increased excitability of cortical neurons^{28,29}.

Finally, VNS/multiple tone pairing also reversed the increase in cortical synchronization observed in noise-exposed rats, but did not reverse the increase in cortical spontaneous activity observed in noise-exposed rats (Fig. 4g, i). There was a trend for the degree of synchronization to be correlated with the degree of gap impairment, but no correlation between the rate of spontaneous activity and the degree of gap impairment (Fig. 4h, j). Our observation that noise-induced increases in spontaneous activity and synchronization are not significantly correlated with behavioural correlates of tinnitus in individual rats is consistent with earlier reports^{19,23}. However, given the potential for small changes in anaesthesia level to influence spontaneous activity and synchronization in the cortex, it remains a possibility that these factors contribute to tinnitus.

Hearing loss, hyperacusis and tinnitus often result from noise exposure and could contribute to the gap impairments observed in this study. Our results confirm that exposure to intense, high-frequency noise causes pathological plasticity that is well correlated with the inability to detect a gap in a mid-frequency, 65-dB SPL tone. Correlations alone do not suggest that these changes cause tinnitus because another confounding factor (such as variability in the degree of cochlear trauma) could cause both variables to be correlated without a causal connection. By randomizing the treatment of rats with identical noise exposure, we were able to eliminate the potential confound caused by variability in the response to noise exposure. Thus, our observation that pairing multiple tone frequencies with VNS can reverse both the neural and behavioural correlates of tinnitus provides good evidence that abnormal activity in the central auditory system is responsible for the subjective experience of tinnitus. In addition, neural correlates of hearing loss (tone thresholds) and hyperacusis (rate level functions) were not correlated with gap impairment in the rats tested (Supplementary Information). Thus, it is reasonable to conclude that the gap impairments observed in this study are primarily related to tinnitus.

VNS-directed plasticity represents a potentially powerful approach to treating tinnitus. Unlike pharmaceutical approaches, this method provides the possibility of generating long-lasting and stimulus-specific changes to neural circuits with minimal side effects. Our control experiments demonstrate that VNS-directed plasticity is driven by the repeated association of VNS with tones, and not by VNS alone. Additional studies are needed to determine whether the pairing of other sensory events with brief periods of VNS could be used to reverse the pathological plasticity associated with other common neurological conditions, such as chronic pain and amblyopia.

METHODS SUMMARY

The VNS–tone pairing protocols, noise exposure procedure, gap detection testing, neurophysiology techniques and analysis are described in Methods. The noise exposure procedure, gap detection testing, and neurophysiology techniques were identical to those in earlier reports^{8,17,25}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 June; accepted 8 November 2010.

Published online 12 January 2011.

1. Flor, H. Phantom-limb pain as a perceptual correlate of cortical reorganization following arm amputation. *Nature* **375**, 482–484 (1995).
2. Eggermont, J. J. & Roberts, L. E. The neuroscience of tinnitus. *Trends Neurosci.* **27**, 676–682 (2004).
3. Möller, A. R. Tinnitus and pain. *Prog. Brain Res.* **166**, 47–53 (2007).
4. Mühlhölzer, W., Elbert, T., Taub, E. & Flor, H. Reorganization of auditory cortex in tinnitus. *Proc. Natl Acad. Sci. USA* **95**, 10340–10343 (1998).

5. De Ridder, D., De Mulder, D., Menovsky, T., Sunaert, S. & Kovacs, S. Electrical stimulation of auditory and somatosensory cortices for treatment of tinnitus and pain. *Prog. Brain Res.* **166**, 377–388 (2007).
6. Okamoto, H., Stracke, H., Stoll, W. & Pantev, C. Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. *Proc. Natl Acad. Sci. USA* **107**, 1207–1210 (2010).
7. Flor, H., Denke, C., Schaefer, M. & Grüsser, S. Effect of sensory discrimination training on cortical reorganization and phantom limb pain. *Lancet* **357**, 1763–1764 (2001).
8. Kilgard, M. P. & Merzenich, M. M. Cortical map reorganization enabled by nucleus basalis activity. *Science* **279**, 1714–1718 (1998).
9. Dorr, A. E. & Debonnel, G. Effect of vagus nerve stimulation on serotonergic and noradrenergic transmission. *J. Pharmacol. Exp. Ther.* **318**, 890–898 (2006).
10. Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A. & Jensen, R. A. Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nature Neurosci.* **2**, 94–98 (1999).
11. Bao, S., Chan, V. T. & Merzenich, N. M. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* **412**, 79–83 (2001).
12. Bollinger, J. J. *Adult Auditory Cortical Plasticity Modulated by Locus Coeruleus Activity*. PhD thesis, Univ. California San Francisco (2006).
13. Ben-Menachem, E. Vagus nerve stimulation, side effects, and long-term safety. *J. Clin. Neurophysiol.* **18**, 415–418 (2001).
14. Noreña, A. J., Tomita, M. & Eggermont, J. J. Neural changes in cat auditory cortex after a transient pure-tone trauma. *J. Neurophysiol.* **90**, 2387–2401 (2003).
15. Salvi, R. J., Wang, J. & Ding, D. Auditory plasticity and hyperactivity following cochlear damage. *Hear. Res.* **147**, 261–274 (2000).
16. Eggermont, J. in *Tinnitus: Pathophysiology and Treatment* (eds Langguth, B., Hajak, G., Kleinjung, T., Cacace, A. & Möller, A. R.) 19–35 (Prog. Brain Res. 166, Elsevier, 2007).
17. Turner, J. G. et al. Gap detection deficits in rats with tinnitus: a potential novel screening tool. *Behav. Neurosci.* **120**, 188–195 (2006).
18. Murphy, W. J. & van Campen, L. E. Temporary threshold shift in ABRs and DPOAEs following noise exposure in Long–Evans rats. *J. Acoust. Soc. Am.* **109**, 2373 (2001).
19. Bauer, C. A., Turner, J. G., Caspary, D. M., Myers, K. S. & Brozoski, T. J. Tinnitus and inferior colliculus activity in chinchillas related to three distinct patterns of cochlear trauma. *J. Neurosci. Res.* **86**, 2564–2578 (2008).
20. Bauer, C. A. & Brozoski, T. J. Assessing tinnitus and prospective tinnitus therapeutics using a psychophysical animal model. *J. Assoc. Res. Otolaryngol.* **2**, 54–64 (2001).
21. Lobarinas, E., Sun, W., Cushing, R. & Salvi, R. A novel behavioral paradigm for assessing tinnitus using schedule-induced polydipsia avoidance conditioning (SIP-AC). *Hear. Res.* **190**, 109–114 (2004).
22. Moore, B. C. & Sandhya, V. The relationship between tinnitus pitch and the edge frequency of the audiogram in individuals with hearing impairment and tonal tinnitus. *Hear. Res.* **261**, 51–56 (2010).
23. Yang, G. et al. Salicylate induced tinnitus: behavioral measures and neural activity in auditory cortex of awake rats. *Hear. Res.* **226**, 244–253 (2007).
24. Noreña, A. J. & Eggermont, J. J. Enriched acoustic environment after noise trauma reduces hearing loss and prevents cortical map reorganization. *J. Neurosci.* **25**, 699–705 (2005).
25. Kilgard, M. P., Vazquez, J. L., Engineer, N. D. & Pandya, P. K. Experience dependent plasticity alters cortical synchronization. *Hear. Res.* **229**, 171–179 (2007).
26. Dietrich, V., Nieschalk, M., Stoll, W., Rajan, R. & Pantev, C. Cortical reorganization in patients with high frequency cochlear hearing loss. *Hear. Res.* **158**, 95–101 (2001).
27. Dauman, R. & Cazals, Y. Auditory frequency selectivity and tinnitus. *Arch. Otorhinolaryngol.* **246**, 252–255 (1989).
28. Kaltenbach, J. A. & Afman, C. E. Hyperactivity in the dorsal cochlear nucleus after intense sound exposure and its resemblance to tone-evoked activity: a physiological model for tinnitus. *Hear. Res.* **140**, 165–172 (2000).
29. Diesch, E., Andermann, M., Flor, H. & Rupp, A. Interaction among the components of multiple auditory steady-state responses: enhancement in tinnitus patients, inhibition in controls. *Neuroscience* **167**, 540–553 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank A. Kuzu, J. Omana, D. Vuppala, H. Rasul, M. Fink, E. Hanack, R. Miller and C. Walker for help with rat behavioural training. We would also like to thank J. Eggermont, A. Möller, C. Bauer, J. Fritz, H. Reed, C. Engineer, A. Reed, M. Brosch, R. Rennaker, R. Beitel, V. Miller, C. McIntyre, G. White, P. Pandya, R. Tyler and D. de Ridder for suggestions about earlier versions of the manuscript. This work was supported by the James S. McDonnell Foundation, the Texas Advanced Research Program, the National Institute for Deafness and other Communication Disorders, and MicroTransponder Inc.

Author Contributions N.D.E., J.R.R., J.D.S., S.P.S. and M.S.B. did the behaviour training sessions, noise exposure and auditory brainstem response recordings. N.D.E., J.R.R., J.D.S., W.A.V. and J.A.S. did cortical microelectrode mappings. J.A.S. did the A1 mapping surgeries. S.P.S. and N.D.E. did all the VNS implant surgeries. M.P.K. and N.D.E. designed the experiments, wrote the manuscript and performed data analysis. All authors discussed the paper and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.D.E. (navzer@utdallas.edu).

METHODS

VNS surgical protocol. Female Sprague–Dawley rats (250–350 g) were implanted with a platinum–iridium bipolar cuff electrode around the left cervical vagus nerve⁹. As in humans, only the left vagus nerve was stimulated because the right vagus nerve contains efferents that stimulate the sinoatrial node and can cause cardiac complications¹³. Leads from the electrode were tunnelled subcutaneously to the top of the head. A four-channel connector was used to deliver current to the stimulating electrode and monitor the electroencephalogram (EEG) during daily VNS sessions. Bone screws placed over the vertex and the cerebellum were used to record auditory brainstem responses (ABRs) and EEG. Each rat was given antibiotics to prevent infection and a single dose of atropine and dexamethazone to reduce fluid accumulations in the lungs immediately after completion of the surgery.

VNS stimulation parameters and single-tone pairing procedures. VNS was delivered to unanaesthetized, unrestrained rats in a $25 \times 25 \times 25$ cm³ wire cage, located inside a $50 \times 60 \times 70$ cm³ chamber lined with acoustic insulating foam. A pilot study was conducted to determine the minimal VNS parameters that reliably reduced EEG amplitude during slow-wave sleep (Supplementary Fig. 3). VNS parameters were identical for every rat in this study. Each 100- μ s, charge-balanced biphasic pulse was delivered with a current of 0.8 mA. The stimulation was delivered as a train of 15 pulses at 30 Hz (500-ms train duration). Cuff impedances were measured daily (~ 5 k Ω). The impedance for three rats was unusually high after implantation and these rats were assigned to the tone-alone and no-therapy groups. The impedance was stable across the duration of training for all other rats. The 500-ms pure tones began 150 ms after the onset of the VNS train (Supplementary Figs 1 and 2). For our earlier nucleus basalis stimulation studies, stimulation beginning either 200 ms before tone onset or 50 ms after tone onset generated indistinguishable map plasticity⁸.

VNS was delivered 300 times per day for 20 days, during a VNS–tone pairing session that lasted 2.5 h (Supplementary Figs 2). To prevent rats from anticipating stimulation timing, there was a 50% chance that VNS would be delivered every 15 s. Twenty-four hours after the last pairing, rats were anaesthetized with pentobarbital and the right auditory cortex was exposed to allow for high-density extracellular microelectrode mapping⁸.

One group of rats ($n = 8$) was exposed to a single 9-kHz, 60-dB SPL tone paired with VNS. No sound was presented when VNS was not delivered. A second group ($n = 5$) was exposed to a 19-kHz, 50-dB SPL tone paired with VNS. During the trials in which no VNS was delivered (50%), a 4-kHz, 50-dB SPL tone was presented. As a result, a 19- or 4-kHz tone was delivered every 15 s. Frequency and intensity calibrations were performed with an ACO Pacific microphone (PS9200-7016) and Tucker-Davis Technologies SIGCAL v4.2 software. The free-field tones were presented from a speaker (Optimus) suspended 20 cm above the wire cage. All paired tones had a 5-ms rise–fall time. The intensity of every tone was selected to be approximately 20 dB SPL above the rat hearing threshold.

Noise exposure and ABRs. Twenty-eight experimental and control rats were barbiturate-anaesthetized and exposed to 16-kHz, 115-dB SPL, octave-band noise for 1 h (refs 17, 20). A single speaker was positioned 5 cm from the left ear. No ear plugs were used to restrict the noise exposure to one ear. Bilateral noise exposure was used because it best approximates the noise exposure that occurs in humans. To confirm cochlear trauma, elevated thresholds were quantified using ABRs in ten rats under pentobarbital anaesthesia before noise exposure, immediately after exposure and 11 weeks after noise exposure (when the auditory cortex was mapped). For ABR recordings, the speaker was positioned 10 cm from the left ear and pure tones (10 ms long, 2.5-ms rise–fall time) were delivered at a rate of 20 Hz. Tone frequencies were 4, 10, 16 and 32 kHz in 10-dB steps from 0 to 85 dB SPL. Tones were randomly interleaved with 1,500 repeats for each frequency–intensity combination. The signals were filtered from 100 to 3,000 Hz and recorded using BRAINWARE v8.12 (Tucker-Davis Technologies). Threshold was defined as the lowest 10-dB SPL step at which an ABR could be recognized (Supplementary Fig. 4).

Gap detection testing. The Turner gap detection method was used to assess a behavioural correlate of tinnitus in every noise-exposed rat¹⁷ (Supplementary Figs 6–8). This method has previously been cross-validated with a conditioned lever suppression task²⁰ ($R = 0.75$) and a licking suppression task²¹. The gap detection method was selected because it avoids the need for food or water deprivation, electric shock or months of behavioural training¹⁷. Testing took place in a $20 \times 20 \times 20$ cm³ wire-mesh cage in a $67 \times 67 \times 67$ cm³ chamber lined with 5-cm acoustic foam. The cage was placed on a startle platform (Lafayette Instrument Co.) that used a piezoelectric transducer to generate a continuous record of downward force. Sounds were generated using System 3 hardware and software (Tucker-Davis Technologies) and were delivered by a speaker (Tucker-Davis Technologies FF1) mounted 20 cm above the cage. Rats underwent gap detection testing with different band-pass-filtered (1,000-Hz bandwidth) sounds centred at 2, 4, 8, 10, 16, 20 and 24 kHz at 65 dB SPL (ref. 17). Startle responses were elicited by a 20-ms burst of white noise at 100 dB SPL. In 50% of trials, a 50-ms gap embedded in the continuous sound served as a warning of a subsequent startling noise and allowed

rats to reduce the amplitude of the response (Supplementary Fig. 7b). The gap in the narrowband noise began 100 ms before the onset of the broadband startling noise. Rats underwent 30 trials during each session. The order of sessions with different continuous sounds was counterbalanced across rats. The interval between each startle sound was 30–35 s.

In untreated noise-exposed rats, gaps in a specific narrowband sound (usually 8 or 10 kHz) did not serve as an effective warning, presumably because the ongoing tinnitus percept prevented the rats from detecting the silent gap. Thus, the animals were not warned that a loud startling noise was coming and exhibited a strong startle response (Supplementary Figs 7b and 8b). Gap detection was quantified as one minus the ratio of the startle amplitude when the startling noise was preceded by a gap in the 65-dB SPL, continuous narrowband sound to the startle amplitude when the startling noise was not preceded by a warning gap. Supplementary Fig. 8 shows typical data from one noise-exposed rat for a session in which the noise burst was cued with a gap in broadband noise (left) and a session in which a gap in an 8-kHz tone served as the warning cue (right). The warning gap typically reduced the startle amplitude by 60–70% (Supplementary Fig. 8a). In noise-exposed rats, gaps in the narrowband noise centred near the low edge of the trauma noise typically reduced the startle amplitude by less than 20%, which is not a statistically significant reduction (Supplementary Fig. 8b). The same procedure was also administered using gaps in 65-dB SPL broadband noise as warning cues of the startling noise (Supplementary Fig. 7a). The frequency with the greatest impairment four weeks after noise exposure is the putative tinnitus frequency for each rat (Fig. 2).

Thirty-six rats were initially tested using the gap startle task for inclusion in this study. Five rats were excluded from the study because they showed no detectable startle response to the noise burst. Of the 31 remaining rats, three were excluded because their startle responses were unusually variable. Twenty-eight rats received noise exposure. Eighteen of these showed a statistically significant impairment in the detection of gaps in one or both mid-frequency (8- or 10-kHz) narrowband sounds tested, relative to gap detection before noise exposure ($P < 0.05$). Three rats were excluded from further study because they no longer showed a startle response to the noise burst (that is, could no longer detect the startle stimulus). Seven rats were excluded from further study because they showed no impairment in gap detection (that is, no evidence of tinnitus). Our observation that gap impairments do not always result from noise exposure is consistent with human and animal studies showing that although hearing loss is common in individuals with tinnitus, the majority of individuals with hearing loss do not have tinnitus^{20,30,31}.

Each of the eighteen rats included in this study showed a significant impairment in its ability to detect a gap in narrowband noise centred on 8 kHz (16 of 18) or 10 kHz (12 of 18). None of the 18 rats showed a significant impairment in the ability to detect a gap in low-frequency narrowband noises (2 or 4 kHz) or in broadband noise (Fig. 2 and Supplementary Fig. 11). This result indicates that these rats are able to respond normally to the startling noise burst and that the mechanisms for modulating the startle response using silent gaps remain intact. Our observation that noise-exposed rats can show gap detection impairments centred at a single frequency or across a narrow range of frequencies is consistent with clinical studies showing significant heterogeneity across subjects in the spectral characteristics of the tinnitus percept^{22,32,33}. Despite this heterogeneity, a large fraction of tinnitus patients can match their tinnitus to a pitch and describe their phantom sound as tonal²².

VNS tone delivery to noise exposed rats. Rats were tested for gap impairment four weeks after noise exposure and 10 and 20 days after the beginning of the sham or experimental therapy. In the VNS/multiple tone paired group ($n = 5$ rats), tones were paired with VNS every 15 s with no VNS–tone pairing 50% of the time. The tone frequencies paired with VNS in the therapy group were designed to reduce the 8–10-kHz region of the frequency map. VNS was repeatedly paired with a 1.3-, 2.2-, 3.7-, 17.8- or 29.9-kHz tone that was randomly selected every trial (300 trials per day). Each tone was presented at ~ 20 dB above the normal hearing threshold for that frequency. The tone-alone control group was passively exposed to the same tones on the same schedule as used in the paired group. A VNS-alone control group received VNS stimulation on the same schedule as used in the paired group without presentation of tones. The third control group did not receive tones or VNS.

To test whether the tinnitus percept remained suppressed after the end of VNS–tone pairing, rats were also tested on gap detection one and three weeks after the end of therapy. At the end of three weeks (that is, 11 weeks after noise exposure), multiunit responses were recorded from auditory cortex neurons from the therapy, sham and naive control rats using dense microelectrode mapping techniques. Physiological and behavioural results from the tone-alone, VNS-alone and no-therapy groups were statistically indistinguishable (Supplementary Fig. 10 and physiological data not shown). Data from the three groups are combined and referred to as sham controls in the main text.

Neurophysiology. In this study, we recorded from a total of 1,492 sites in 21 rats ($n = 8$ naive controls, $n = 5$ VNS therapy and $n = 8$ sham controls). Nine hundred and sixty-five of those sites were in A1 and were included in the analysis presented in this report. We recorded 220 multiunit responses from A1 sites in noise-exposed rats that received VNS/multiple tone pairing ($n = 5$). We also recorded 321 A1 sites from noise-exposed rats that did not receive VNS/tone pairing ($n = 8$). The latter group included noise-exposed rats that received tones with no VNS ($n = 3$), VNS with no tones ($n = 2$) or no therapy ($n = 3$). Because neural and behavioural responses were similar in all three control groups, the results were pooled to form a single data set referred to as the sham therapy group. During the acute electrophysiology recordings, sounds were delivered in a foam-lined, double-walled, sound-attenuated chamber using a speaker (Motorola 40-1221) positioned directly opposite the left ear at a distance of 10 cm. Multiunit responses were recorded using Parylene-coated tungsten electrodes that were glued together (250- μm separation, 2 M Ω at 1 kHz; FHC) and lowered approximately 500 μm below the cortical surface. Frequency and intensity calibrations were performed with an ACO Pacific microphone (PS9200-7016) and Tucker-Davis Technologies SIGCAL v4.2 software. Auditory frequency tuning curves were determined at each site by presenting 81 logarithmically spaced frequencies spanning 1 to 32 kHz at 16 intensities from 0 to 75 dB SPL (1,296 total stimuli). The tones (25-ms duration, 5-ms rise–fall time) were randomly interleaved and separated by 500 ms. Tuning curve parameters were determined by an experienced blind observer using custom software written in MATLAB v7.9 (Mathworks) to randomize the order of data from each recording site across all groups. Experimenters were blind to the experimental conditions of each rat during electrophysiology recordings.

Data analysis. Gap discrimination was quantified as the percentage inhibition of the startle response when a gap (warning cue) preceded the startling noise relative to the startle response when no gap was present¹⁷. Eight of 36 rats tested failed to generate consistent startle responses and were excluded from the study before noise exposure. Noise exposure eliminated the startle response in three of the remaining 28 rats, and these rats were excluded from the study. Noise exposure failed to generate any impairment in gap detection in seven of the remaining 25 rats, and these rats were also excluded from the study. Eighteen noise-exposed rats were included in the study. Neural responses were collected from thirteen rats (five

VNS–tone paired rats and eight sham therapy rats). One rat died before neural responses could be collected. Only behavioural responses (and EEG) were collected from the remaining four rats (two treated rats and two shams) so that the duration of the benefit could be estimated.

Sites were determined to be in A1 on the basis of continuous tonotopy. At each A1 recording site, characteristic frequency, frequency bandwidth, response threshold, spontaneous rate and latency were determined using a standard method in which the experimenter was blind to the experimental group and recording location⁸. At each pair of simultaneously recorded A1 sites, neural synchrony during silence (300 s) was quantified as the cross-correlation function²⁵. The peak in the cross-correlation function (with or without subtraction of the shift predictor) was also computed and gave similar results to Pearson correlation coefficient (R). Map plasticity was quantified as the percent of A1 neurons with a characteristic frequency in a given range or as the percent of A1 neurons responding to each frequency–intensity combination using the Voronoi tessellation method of interpolation^{8,34}. Frequency selectivity was quantified as the bandwidth 10, 20, 30 or 40 dB above threshold. Results were similar regardless of the intensity above threshold used. Excitability was quantified as the number of spikes evoked by each tone within each site's receptive field and as the spontaneous activity rate during silence.

All protocols and recording procedures comply with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee at the University of Texas at Dallas.

30. Bauer, C. A., Brozoski, T. J. & Myers, K. Primary afferent dendrite degeneration as a cause of tinnitus. *J. Neurosci. Res.* **85**, 1489–1498 (2007).
31. König, O., Schaette, R., Kempter, R. & Gross, M. Course of hearing loss and occurrence of tinnitus. *Hear. Res.* **221**, 59–64 (2006).
32. Burns, E. M. A comparison of variability among measurements of subjective tinnitus and objective stimuli. *Audiology* **23**, 426–440 (1984).
33. Ochi, K., Ohashi, T. & Kenmochi, M. Hearing impairment and tinnitus pitch in patients with unilateral tinnitus: comparison of sudden hearing loss and chronic tinnitus. *Laryngoscope* **113**, 427–431 (2003).
34. Engineer, N. D. et al. Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons. *J. Neurophysiol.* **92**, 73–82 (2004).

Quantum storage of photonic entanglement in a crystal

Christoph Clausen^{1*}, Imam Usmani^{1*}, Félix Bussi eres¹, Nicolas Sangouard¹, Mikael Afzelius¹, Hugues de Riedmatten^{1,2,3} & Nicolas Gisin¹

Entanglement is the fundamental characteristic of quantum physics—much experimental effort is devoted to harnessing it between various physical systems. In particular, entanglement between light and material systems is interesting owing to their anticipated respective roles as ‘flying’ and stationary qubits in quantum information technologies (such as quantum repeaters^{1–3} and quantum networks⁴). Here we report the demonstration of entanglement between a photon at a telecommunication wavelength (1,338 nm) and a single collective atomic excitation stored in a crystal. One photon from an energy–time entangled pair⁵ is mapped onto the crystal and then released into a well-defined spatial mode after a predetermined storage time. The other (telecommunication wavelength) photon is sent directly through a 50-metre fibre link to an analyser. Successful storage of entanglement in the crystal is proved by a violation of the Clauser–Horne–Shimony–Holt inequality⁶ by almost three standard deviations ($S = 2.64 \pm 0.23$). These results represent an important step towards quantum communication technologies based on solid-state devices. In particular, our resources pave the way for building multiplexed quantum repeaters⁷ for long-distance quantum networks.

Although single atoms^{8,9} and cold atomic gases^{10–15} are currently some of the most advanced light–matter quantum interfaces, there is a strong motivation to control light–matter entanglement with more practical systems, such as solid-state devices¹⁶. Solid-state quantum memories for photons can be implemented with cryogenically cooled crystals doped with rare-earth-metal ions¹⁷, which have impressive coherence properties at temperatures below 4 K. They have the advantage of simple implementation because rare-earth-metal-doped crystals are widely produced for solid-state lasers, and closed-cycle cryogenic coolers are commercially available. Important progress has been made over the last years in the context of light storage into solid-state memories, including long storage times¹⁸, high efficiency¹⁹ and storage of light at the single photon level with high coherence and negligible noise^{19–23}. Yet these experiments were realized with classical bright pulses or weak coherent states of light. Although this is sufficient to characterize the performance of the memory, and even to infer the quantum characteristics of the device^{19,20}, it is not sufficient for the implementation of more sophisticated experiments involving entanglement, as required for most applications in quantum information science. For this purpose, it is necessary to store non-classical light, in particular individual photons that are part of an entangled state (generated, for example, through spontaneous parametric down-conversion, SPDC), similar to previous demonstrations using electromagnetically induced transparency in cold atomic gases^{14,15}. In addition, for quantum communication applications, the other part of the entangled state should be a photon at telecommunication wavelength in order to minimize loss during transmission in optical fibres.

In this Letter, we report on an experiment in which a photon from an entangled pair is stored in a quantum memory based on a rare-earth-metal-doped crystal. More specifically, we show that non-classical

intensity correlations between the two photons still exist after storage and retrieval. We then show, through a violation of a Bell inequality, that the storage process creates a light–matter entangled state. In addition, these results represent the first successful mapping of energy–time entangled photons onto a quantum memory.

Our experiment consists of a coherent solid-state quantum memory and a source of entangled photons. A schematic of the experiment is shown in Fig. 1. The source is based on non-degenerate SPDC in a nonlinear waveguide pumped by continuous wave light at 532 nm. This yields energy–time entangled photons with the signal photon at

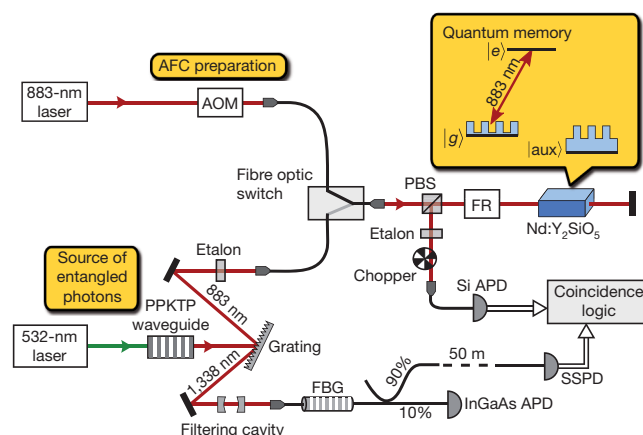


Figure 1 | Experimental set-up. The experimental set-up can be divided into three parts: the Nd:Y₂SiO₅ crystal serving as quantum memory, the laser system for the preparation of the AFC in the crystal, and the source of entangled photons with associated spectral filtering. During the experiment we periodically switch between 15 ms of AFC preparation and frequency stabilization and a 15-ms measurement phase, in which single photons are stored. During the preparation, the comb structure is prepared by frequency-selective optical pumping of atoms from the ground state $|g\rangle$ to the auxiliary state $|aux\rangle$ using light from an 883-nm diode laser in combination with an acousto-optic modulator (AOM). The fibre optic switch is in the upper position, and the silicon avalanche photodiode (Si APD) is protected from the bright light by a chopper. During the measurement phase, the positions of switch and chopper are reversed. Now, photon pairs are generated in the periodically poled potassium titanyl phosphate (PPKTP) waveguide via SPDC. The two photons in a pair are spatially separated by a diffraction grating and then strongly filtered by two etalons, a cavity and a fibre Bragg grating (FBG). Photons at 883 nm are sent through the crystal in a double-pass configuration to increase the absorption probability, and are afterwards detected by the Si avalanche photodiode. Photons at 1,338 nm are directed towards a superconducting single photon detector (SSPD) located in another laboratory 50 m away. All relevant quantities are extracted from the coincidence statistics of the two detectors. Details of the frequency stabilization and the filtering system are given in the Methods. PBS, polarizing beam splitter; FR, Faraday rotator.

¹Group of Applied Physics, University of Geneva, CH-1211 Geneva 4, Switzerland. ²ICFO—Institut de Ci ncies Fot niques, Mediterranean Technology Park, 08860 Castelldefels, Barcelona, Spain. ³ICREA—Institut Catal  de Recerca i Estudis Avan ats, 08015 Barcelona, Spain.

*These authors contributed equally to this work.

the memory wavelength of 883 nm, and the idler photon at the telecom wavelength of 1,338 nm. Both photons initially have a spectral width of approximately 1.5 THz, a factor of 10^4 larger than the 120-MHz bandwidth of the memory. Hence, strong filtering is crucial¹⁴ to achieve signal-to-noise ratios sufficiently large to reveal the presence of entanglement during storage. After filtering, the signal photon is sent to the memory, and the idler photon is coupled into a fibre leading to a detector in another laboratory 50 m away. Owing to the low loss at telecommunication wavelengths, this distance could, in principle, be extended to several kilometres without significantly affecting the results presented here.

The quantum memory is a 1-cm-long Y_2SiO_5 crystal, impurity-doped with neodymium ions having a resonance at 883 nm with good coherence properties²³. It is based on a photon-echo-type interaction using an atomic frequency comb (AFC) (see ref. 24 and Supplementary Information). In an AFC, the absorption profile of the atomic ensemble is shaped into a comb-like structure by optical pumping. A photon is then, with some efficiency, absorbed and re-emitted into a well-defined spatial mode due to a collective rephasing of the atoms in the comb structure. The time of re-emission depends on the period of the comb and is predetermined. We have previously shown that this kind of memory can store multiple temporal modes²³ and is therefore perfectly suited for storing energy-time entangled photons. For the work presented here, we have significantly improved the storage efficiency to obtain sufficiently large signal-to-noise ratios. Indeed, using a new optical pumping scheme for the preparation of the AFC (see Supplementary Information), the efficiency was increased by a factor of three for storage times below 200 ns, now reaching values up to 21% (see results below).

In a first experiment we verified that the non-classical nature of the intensity correlations between the signal (883 nm) and idler (1,338 nm) modes is preserved after the storage and retrieval process. If we assume second-order auto-correlations of signal and idler $g_x^{(2)}$ (where $x = 's'$ for signal or $'i'$ for idler) satisfying $1 \leq g_x^{(2)} \leq 2$, then non-classicality is proved by measuring a cross-correlation $g_{si}^{(2)} = P_{si}/P_s P_i$ greater than 2 (see ref. 25). Here, P_s (or P_i) is the probability of detecting a signal (or idler) photon, and P_{si} the probability of a coincidence detection (see Methods).

We first measured $g_{si}^{(2)}$ as a function of the pump power of the source, as shown in Fig. 2a. We find an optimum around a pump power of 3 mW, where $g_{si}^{(2)} \approx 115$ without the AFC memory, and $g_{si}^{(2)} \approx 30$ after a 25-ns storage, thus proving the quantum character of the storage (note that all results presented in this Letter are without any subtraction of background noise). The reduction in the cross-correlation with

the storage is due to limited efficiency (21%), which effectively increases the contribution of accidental coincidences stemming from dark counts and multiple pair emissions. Next, we measured the memory efficiency and the cross-correlation for different storage times, as shown in Fig. 2b and c.

We now turn our attention towards a particular kind of quantum correlation, namely entanglement. By performing a two-photon quantum interference experiment, we show that the entanglement of the photon pair is preserved when the signal photon is stored in the crystal.

Photon pairs generated by our source are energy-time entangled, that is, the two photons in a pair are created simultaneously to ensure energy conservation, but the pair-creation time is uncertain to within the coherence time of the pump laser. We wish to reveal the presence of this entanglement using a Franson-type set-up⁵. As detailed in the Supplementary Information, the correlations can be interpreted as stemming from local measurements performed on a post-selected time-bin entangled state: $\frac{1}{\sqrt{2}}(|E_s E_i\rangle + |L_s L_i\rangle)$, where the early and late time bins $|E_{s,i}\rangle$ and $|L_{s,i}\rangle$ are separated by a time of 25 ns set by the analysing interferometer (see Fig. 3a). In our experiment, however, the state of the signal photon is stored as a collective atomic excitation in the quantum memory before the measurement. Moreover, using a double AFC scheme^{20,23}, the memory is used not only to store the entangled photon, but also to analyse it as part of the measurement. More precisely, the incident time-bins $|E_s\rangle$ and $|L_s\rangle$ are mapped to distinct AFC modes $|E_{QM}\rangle$ and $|L_{QM}\rangle$, respectively (where subscript QM denotes quantum memory). Storage of the entangled signal photon then creates a light-matter entangled state:

$$\frac{1}{\sqrt{2}}(|E_{QM} E_i\rangle + |L_{QM} L_i\rangle) \quad (1)$$

The predetermined storage times of $|E_{QM}\rangle$ and $|L_{QM}\rangle$ are 75 ns and 50 ns, respectively. After absorption, both AFCs coherently re-emit the stored excitation into the same well-defined temporal and spatial mode with a relative phase $\Delta\phi_s$. This re-emission, followed by detection, constitutes the measurement of the state of the memory. The idler photon is measured using a fibre-based time-bin qubit analyser with a 25-ns delay and a relative phase $\Delta\phi_i$ between the short and long arms. The coincidence detection probability is given by:

$$P_{si} \propto 1 + V \cos(\Delta\phi_s + \Delta\phi_i) \quad (2)$$

where V is the visibility of interference. Figure 3b shows the measured coincidence rate as a function of $\Delta\phi_s$ for two values of $\Delta\phi_i$. The raw visibilities are $V = (84 \pm 4)\%$ and $(78 \pm 4)\%$.

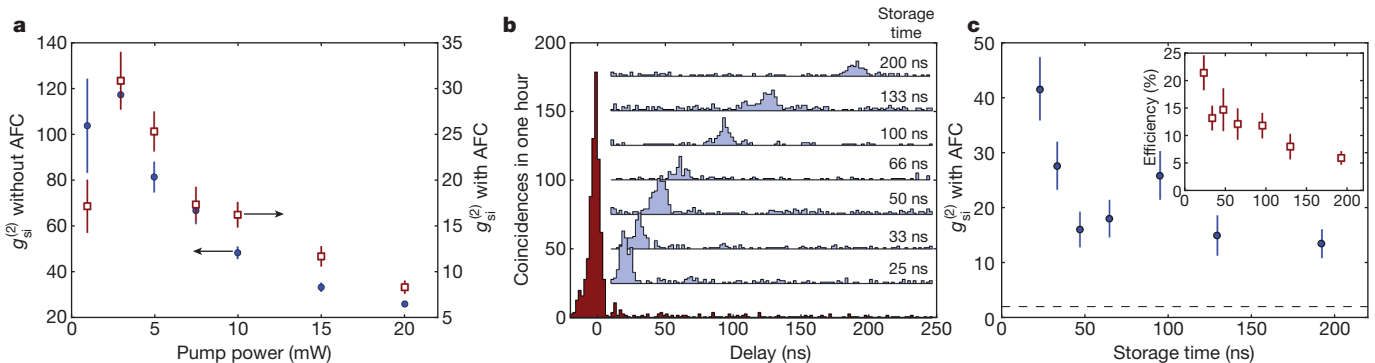


Figure 2 | Non-classical correlations and storage efficiency. **a**, Cross-correlation $g_{si}^{(2)}$ as a function of the pump power incident on the wave guide. Data points shown were taken with an AFC memory storage time of 25 ns (brown square symbols), and for comparison, with the crystal prepared with a 120-MHz-wide transmission window, that is, without AFC (blue circle symbols). The size of the coincidence window is 10 ns. **b**, Coincidence histograms for different predetermined storage times, vertically offset for clarity. For comparison, the lowest histogram was taken without AFC. The pump power was 3 mW. **c**, Cross-

correlation $g_{si}^{(2)}$ as a function of storage time with 10-ns coincidence window, extracted from **b**. For storage times up to 200 ns the correlations stay well above the classical limit of $g_{si}^{(2)} = 2$ (dashed line). The inset shows the storage efficiency for the same range of storage times. With increasing storage times, limiting factors in the storage medium degrade the comb shape and reduce the efficiency and cross-correlation (see Supplementary Information). However, the latter stays well above the classical limit for storage times up to 200 ns. Error bars show ± 1 standard deviation (s.d.).

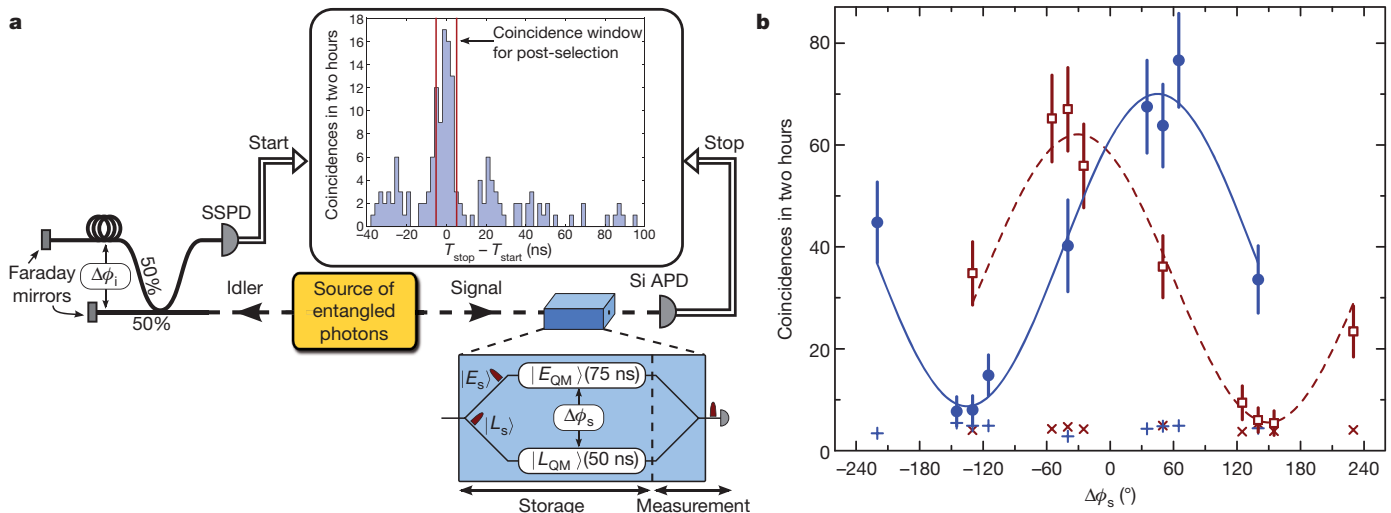


Figure 3 | Storage of photonic entanglement in a crystal. **a**, Franson-type set-up used to reveal the entanglement. A qubit analyser consisting of an unbalanced, fibre-based Michelson interferometer with 25-ns delay and relative phase $\Delta\phi_i$ is inserted before the SSPD used to detect the idler photon (see also Fig. 1). The signal photon is stored in the crystal, yielding a light-matter entangled state. The state of the memory is measured through re-emission and detection of the photon in the time-window at zero-time delay (central peak) of the coincidence histogram (inset). This post-selects measurement on the entangled state of equation (1). The relative phase $\Delta\phi_s$ can

Quantum entanglement can be revealed by a violation of the Clauser–Horne–Shimony–Holt (CHSH) inequality⁶. The possibility of violating this inequality, that is, of finding a CHSH parameter $S > 2$, can be inferred indirectly from a visibility larger than $1/\sqrt{2} \approx 70.7\%$. Nevertheless, we performed the measurements necessary for a direct violation of the inequality and obtained $S = 2.64 \pm 0.23$. This proves the presence of entanglement between the idler photon and the matter qubit in the crystal, provided the effect of the memory on single photons is appropriately described as storage followed by measurement (see Supplementary Information). This description is correct within the theory of AFC memories²⁴, which is supported by experiments storing weak coherent states of light^{11,21–23}. Note also that we do not claim any demonstration of nonlocal correlations. Indeed, besides the usual locality and detection loopholes, here the measurement setting has to be chosen before the photonic qubit is mapped onto the crystal. This could have been avoided by adding an interferometer after the memory, the latter being used for storage only. We did not do so because we think that it is elegant and simple to use the memory also as a small quantum processor that performs the measurement.

A particularly intriguing situation arises when post-selecting on the case where only $|E_s\rangle$ is stored in the crystal for 25 ns using a single AFC scheme, while $|L_s\rangle$ is directly transmitted. Indeed, the imbalance between the storage efficiency and the transmission probability offers a well-suited qubit analyser for a violation of the CHSH inequality using bases lying in the x - z plane of the Bloch sphere. We performed such a measurement and observed $S = 2.62 \pm 0.15$ (see Supplementary Information). This implies that the initial photon–photon entangled state is mapped onto a state of the form:

$$\sqrt{\eta_{\text{abs}}} |E_{\text{QM}} E_i\rangle + |L_s L_i\rangle \quad (3)$$

where η_{abs} is the absorption efficiency. This is an entangled state between a telecommunication-wavelength qubit and a light-matter hybrid qubit. We note that this kind of hybrid qubit is the key ingredient of an efficient quantum repeater protocol based on atomic ensembles and linear optics³.

This work is part of the effort towards implementing a quantum repeater, which could provide a solution to the distance limit (due to intrinsic loss) for entanglement distribution and quantum cryptography

be reliably set to any desired value (see the Supplementary information).

b, Number of coincidences in the central peak in two hours as a function of the relative phase $\Delta\phi_s$ for two values of $\Delta\phi_i$. The pump power was 5 mW, and the size of the coincidence window 10 ns. The solid and dashed lines result from fits to equation (2) and respectively give visibilities of $V = (78 \pm 4)\%$ and $(84 \pm 4)\%$. The visibilities are mainly limited by the level of accidental coincidences (cross symbols). The fit also gives a difference between the two values of $\Delta\phi_i$ of $75^\circ \pm 10^\circ$. These values closely match settings necessary for a maximal violation of the CHSH inequality. Error bars are ± 1 s.d.

using optical fibres³. To achieve this long-term goal, several future advances are required. The user must be able to trigger the re-emission of the memory, whereas in our experiment the duration of the storage is pre-determined. We have proposed²⁴ and demonstrated²⁶ a method for achieving on-demand re-emission using so-called spin-wave storage. This has the additional benefit of allowing longer storage times owing to the more robust spin coherence. Another crucial aspect is the efficiency, which is directly linked to the optical depth of the material²⁴. It can be increased by using longer crystals¹⁹ or optical cavities^{27,28}.

The creation of entanglement between a single photon and a macroscopic object—in this case a single collective atomic excitation delocalized over a 1-cm-long crystal—is fascinating in itself. Beyond its fundamental interest, we believe that our demonstration of storage of entanglement in a crystal represents an important step towards quantum repeaters based on solid-state quantum memories.

We note that, parallel to this work, Saglamyurek *et al.* have demonstrated storage and retrieval of an entangled photon using a thulium-doped lithium niobate waveguide²⁹.

METHODS SUMMARY

Spectral filtering and detection. The bandwidth of the photon pairs is reduced by a factor of 10^4 in several steps. Pump, signal and idler photons are spatially separated by a diffraction grating (see Fig. 1). In combination with coupling into single-mode fibres, this reduces the bandwidth to tens of gigahertz. A subsequent passage through a Fabry–Perot cavity reduces the bandwidth of the idler photon to 45 MHz (corresponding to a coherence time of about 4 ns), and a fibre Bragg grating blocks all but one of the longitudinal cavity modes. The signal photon is filtered by two etalons with a linewidth of 600 MHz each, and different free spectral ranges. The detector efficiency is 8% for the idler photon with 10-Hz dark counts, and 30% with 100-Hz dark counts for the signal photon.

Frequency stabilization. We must ensure, for the whole duration of a measurement, that the central frequency of the optical filtering system at 1,338 nm and of the AFC at 883 nm both satisfy the energy conservation of the SPDC process. To do this, a small fraction of the light at 883 nm is overlapped with the light of the 532-nm laser that pumps the PPKTP waveguide. This leads to the creation of light at 1,338 nm by difference frequency generation (DFG). Using this DFG signal, the frequency of the 532-nm light is adjusted such that the detection rate on a separate InGaAs avalanche photodiode (see Fig. 1) stays constant, which means that the 1,338-nm DFG light is in resonance with the filtering cavity. Long-term stability of the 883-nm laser itself is achieved by continuously referencing it to a Fabry–Perot cavity.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 August; accepted 9 November 2010.

Published online 12 January 2011.

1. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
2. Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
3. Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. Preprint at (<http://arxiv.org/abs/0906.2699>) (2009).
4. Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
5. Franson, J. D. Bell inequality for position and time. *Phys. Rev. Lett.* **62**, 2205–2208 (1989).
6. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
7. Simon, C. *et al.* Quantum repeaters with photon pair sources and multimode memories. *Phys. Rev. Lett.* **98**, 190503 (2007).
8. Blinov, B. B., Moehring, D., Duan, L.-M. & Monroe, C. Observation of entanglement between a single trapped ion and a single photon. *Nature* **428**, 153–157 (2004).
9. Volz, J. *et al.* Observation of entanglement of a single photon with a trapped atom. *Phys. Rev. Lett.* **96**, 030404 (2006).
10. Matsukevich, D. N. *et al.* Entanglement of a photon and a collective atomic excitation. *Phys. Rev. Lett.* **95**, 040405 (2005).
11. de Riedmatten, H. *et al.* Direct measurement of decoherence for entanglement between a photon and stored atomic excitation. *Phys. Rev. Lett.* **97**, 113603 (2006).
12. Chen, S. *et al.* Demonstration of a stable atom-photon entanglement source for quantum repeaters. *Phys. Rev. Lett.* **99**, 180505 (2007).
13. Sherson, J. F. *et al.* Quantum teleportation between light and matter. *Nature* **443**, 557–560 (2006).
14. Akiba, K., Kashiwagi, K., Arikawa, M. & Kozuma, M. Storage and retrieval of nonclassical photon pairs and conditional single photons generated by the parametric down-conversion process. *N. J. Phys.* **11**, 013049 (2009).
15. Jin, X.-M. *et al.* Quantum interface between frequency-uncorrelated down-converted entanglement and atomic-ensemble quantum memory. Preprint at (<http://arxiv.org/abs/1004.4691>) (2010).
16. Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
17. Tittel, W. *et al.* Photon-echo quantum memory in solid state systems. *Laser Photon. Rev.* **4**, 244–267 (2010).
18. Longdell, J. J., Fraval, E., Sellars, M. J. & Manson, N. B. Stopped light with storage times greater than one second using electromagnetically induced transparency in a solid. *Phys. Rev. Lett.* **95**, 063601 (2005).
19. Hedges, M. P., Longdell, J. J., Li, Y. & Sellars, M. J. Efficient quantum memory for light. *Nature* **465**, 1052–1056 (2010).
20. de Riedmatten, H., Afzelius, M., Staudt, M. U., Simon, C. & Gisin, N. A solid-state light-matter interface at the single-photon level. *Nature* **456**, 773–777 (2008).
21. Chanelière, T., Ruggiero, J., Bonarota, M., Afzelius, M. & Gouët, J.-L. L. Efficient light storage in a crystal using an atomic frequency comb. *N. J. Phys.* **12**, 023025 (2010).
22. Sabooni, M. *et al.* Storage and recall of weak coherent optical pulses with an efficiency of 25%. *Phys. Rev. Lett.* **105**, 060501 (2010).
23. Usmani, I., Afzelius, M., de Riedmatten, H. & Gisin, N. Mapping multiple photonic qubits into and out of one solid-state atomic ensemble. *Nature Commun.* **1**, 12 (2010).
24. Afzelius, M., Simon, C., de Riedmatten, H. & Gisin, N. Multimode quantum memory based on atomic frequency combs. *Phys. Rev. A* **79**, 052329 (2009).
25. Kuzmich, A. *et al.* Generation of nonclassical photon pairs for scalable quantum communication with atomic ensembles. *Nature* **423**, 731–734 (2003).
26. Afzelius, M. *et al.* Demonstration of atomic frequency comb memory for light with spin-wave storage. *Phys. Rev. Lett.* **104**, 040503 (2010).
27. Afzelius, M. & Simon, C. Impedance-matched cavity quantum memory. *Phys. Rev. A* **82**, 022310 (2010).
28. Moiseev, S. A., Andrianov, S. N. & Gubaidullin, F. F. Efficient multimode quantum memory based on photon echo in an optimal QED cavity. *Phys. Rev. A* **82**, 022311 (2010).
29. Saglamyurek, E. *et al.* Broadband waveguide quantum memory for entangled photons. *Nature* doi:10.1038/nature09719 (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Locher for help during the early stages of the experiment. We are grateful to A. Beveratos and W. Tittel for lending us avalanche photodiodes. This work was supported by the Swiss NCCR Quantum Photonics, the Science and Technology Cooperation Program Switzerland–Russia, as well as by the European projects QuRep and ERC-Qore. F.B. was supported in part by FQRNT.

Author Contributions All authors contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A. (mikael.afzelius@unige.ch).

METHODS

Spectral filtering and detection. The narrowband filtering of the SPDC photons consists of several steps (see Fig. 1). First, a diffraction grating spatially separates the pump, signal and idler photons and, in combination with coupling into single-mode fibres, reduces the bandwidth of the photons at 883 nm (or 1,338 nm) to 90 GHz (or 60 GHz). Photons at 1,338 nm are then coupled through a Fabry–Perot cavity with linewidth 45 MHz and free spectral range of 23.9 GHz. Subsequently, a fibre Bragg grating with 16 GHz bandwidth ensures that only a single longitudinal cavity mode remains.

Filtering one of the photons in the pair is the same as filtering the photon pair as a whole, because energy conservation guarantees that photons measured in coincidence have the same bandwidth. However, uncorrelated photons would then contribute significantly to the accidental coincidences. Therefore, complementary filtering at 883 nm was necessary. To do this, we used one solid and one air-spaced etalon, both with bandwidths around 600 MHz. Different free spectral ranges of 42 and 50 GHz eliminate uncorrelated longitudinal modes. Additionally, outside the 120-MHz bandwidth of the AFC, the absorption of the crystal with an inhomogeneous linewidth of 6 GHz provides a final filtering step.

We used detectors with 30% detection efficiency and approximately 100 Hz dark counts at 883 nm, and detectors with 8% detection efficiency and approximately 10 Hz at 1,338 nm. Together with a transmission of the filtering system for the signal (or idler) photon of 45% (or 14%), and 4% (or 14%) for the remainder of the optical set-up, we reached an overall detection efficiency of 0.5% (or 0.15%) (see also Supplementary Information). These numbers could, in principle, be significantly improved through optimized optical alignment, the use of anti-reflection-coated elements, and so on.

Frequency stabilization. In the experiment, coincidence rates are typically a few per minute. With accumulation times thus reaching several hours, a high degree of frequency stability of the lasers and filtering elements is indispensable. In particular, frequency drifts of the AFC preparation laser with respect to the pump laser of the

SPDC source have to be eliminated. Otherwise, the photon-pair frequencies $\omega_{883} + \omega_{1338} = \omega_{532}$ imposed by energy conservation in the SPDC would not simultaneously match the centre of the AFC and that of the filtering system at 1,338 nm. Drifts were eliminated using the following method. First, the long-term stability of the 883-nm laser was dramatically increased by locking it to a temperature-stabilized Fabry–Perot cavity. Second, during the 15-ms preparation cycle, we injected a fraction of the 883-nm light into the waveguide. The frequency of the light created at 1,338 nm via difference frequency generation (DFG) was tuned by controlling the frequency of the pump laser at 532 nm. Using a side-of-fringe technique, we could then lock the frequency of the DFG signal to the transmission peak of the filtering cavity. As a result, long-term frequency deviations between the centre of the AFC structure and the filtered photon pairs were reduced to about 1 MHz over several hours.

For measurements involving the unbalanced Michelson interferometer for the idler photon, the phase of the interferometer was also stabilized using the highly coherent DFG light.

Photon correlations in SPDC. Neglecting the exact frequency dependence, the state of the photons created in the SPDC process is described by $|0_s, 0_i\rangle + \sqrt{p}|1_s, 1_i\rangle + O(p)$, where the subscript 's' (or 'i') indicates the signal (or idler) mode at 883 nm (or 1,338 nm). Here, the pair creation probability p is assumed to be small and proportional to the pump power. In such a state, the signal and idler modes individually exhibit the statistics of a classical thermal field, that is, their auto-correlations are $g_x^{(2)} = 2$ for $x = s$ or i . We stress, however, that the criterion for non-classicality of the cross-correlation that we used, namely $g_{si}^{(2)} = P_{si}/P_s P_i > 2$, requires only that $1 \leq g_x^{(2)} \leq 2$, which is always fulfilled by non-degenerate photon pairs created through SPDC. In practice, P_{si} (or $P_s P_i$) is determined by the number of coincidences in a certain time window centred on (or away from) the coincidence peak. For low pump powers, the measured cross-correlation is usually limited by detector dark counts, and at high pump powers it is reduced by the contribution of multiple pairs, that is, higher-order terms in p .

A massive protocluster of galaxies at a redshift of $z \approx 5.3$

Peter L. Capak¹, Dominik Riechers², Nick Z. Scoville², Chris Carilli³, Pierre Cox⁴, Roberto Neri⁴, Brant Robertson², Mara Salvato⁵, Eva Schinnerer⁶, Lin Yan¹, Grant W. Wilson⁷, Min Yun⁷, Francesca Civano⁸, Martin Elvis⁸, Alexander Karim⁶, Bahram Mobasher⁹ & Johannes G. Staguhn¹⁰

Massive clusters of galaxies have been found that date from as early as 3.9 billion years¹ (3.9 Gyr; $z = 1.62$) after the Big Bang, containing stars that formed at even earlier epochs^{2,3}. Cosmological simulations using the current cold dark matter model predict that these systems should descend from ‘protoclusters’—early overdensities of massive galaxies that merge hierarchically to form a cluster^{4,5}. These protocluster regions themselves are built up hierarchically and so are expected to contain extremely massive galaxies that can be observed as luminous quasars and starbursts^{4–6}. Observational evidence for this picture, however, is sparse because high-redshift protoclusters are rare and difficult to observe^{6,7}. Here we report a protocluster region that dates from 1 Gyr ($z = 5.3$) after the Big Bang. This cluster of massive galaxies extends over more than 13 megaparsecs and contains a luminous quasar as well as a system rich in molecular gas⁸. These massive galaxies place a lower limit of more than 4×10^{11} solar masses of dark and luminous matter in this region, consistent with that expected from cosmological simulations for the earliest galaxy clusters^{4,5,7}.

Cosmological simulations predict that the progenitors of present-day galaxy clusters are the largest structures at high redshift^{4,5,7} ($M_{\text{halo}} > 2 \times 10^{11}$ solar masses (M_{\odot}) and $M_{\text{stars}} > 4 \times 10^9 M_{\odot}$ at $z \approx 6$). These protocluster regions should be characterized by local overdensities of massive galaxies on co-moving distance scales of 2–8 Mpc that coherently extend over tens of megaparsecs, forming a structure that will eventually coalesce into a cluster^{4,5,7,9}. Furthermore, owing to the high mass densities and correspondingly high merger rates, extreme phenomena such as starbursts and quasars should preferentially exist in these regions^{4–7,9,10}. Although overdensities have been reported around radio galaxies on ~ 10 –20-Mpc scales^{6,7} and large gas masses around quasars^{11,12} at redshifts greater than $z = 5$, the available data is not comprehensive enough to constrain the mass of these protoclusters and hence provide robust constraints on cosmological models^{6,7,9}.

We used data covering the entire accessible electromagnetic spectrum in the 2-square-degree Cosmological Evolution Survey (COSMOS) field¹³ (right ascension, 10 h 00 min 30 s; declination, $2^{\circ} 30' 00''$) to search for starbursts, quasars and massive galaxies as signposts of potential overdensities at high redshift. This deep, large-area field provides the multiwavelength data required to find protoclusters on scales > 10 Mpc ($5'$). Optically bright objects at redshifts greater than $z = 4$ were identified through optical and near-infrared colours. Extreme star formation activity was found using millimetre-wave^{14,15} and radio¹⁶ measurements, and potential luminous quasars were identified by X-ray measurements¹⁷. Finally, extreme objects and their surrounding galaxies were targeted with the Keck II telescope and the Deep Extragalactic Imaging

Multi-Object Spectrograph (W. M. Keck Observatory, Hawaii) to measure redshifts.

We found a grouping of four major objects at $z = 5.30$ (Fig. 1). The most significant overdensity appears near the extreme starburst galaxy COSMOS AzTEC-3, which contains $> 5.3 \times 10^{10} M_{\odot}$ of molecular gas and has a dynamical mass, including dark matter, of $> 1.4 \times 10^{11} M_{\odot}$ (ref. 8). The far-infrared (60–120- μm) luminosity of this system is estimated to be $(1.7 \pm 0.8) \times 10^{13}$ solar luminosities (L_{\odot}), corresponding to a star formation rate of $> 1,500 M_{\odot}$ per year¹⁸, which is > 100 times the rate of an average galaxy (with luminosity L_{*}) at $z = 5.3$ (ref. 19). The value and error given are the mean estimate and scatter derived from empirical estimates based on the submillimetre flux, radio flux limit, and CO luminosity, along with model fitting. The models predict a much broader range in total infrared (8–1,000- μm) luminosities, ranging from $2.2 \times 10^{13} L_{\odot}$ to $11 \times 10^{13} L_{\odot}$. The large uncertainty results from the many assumptions used in the models, combined with a lack of data constraining the infrared emission at wavelengths less than rest-frame 140 μm . However, the

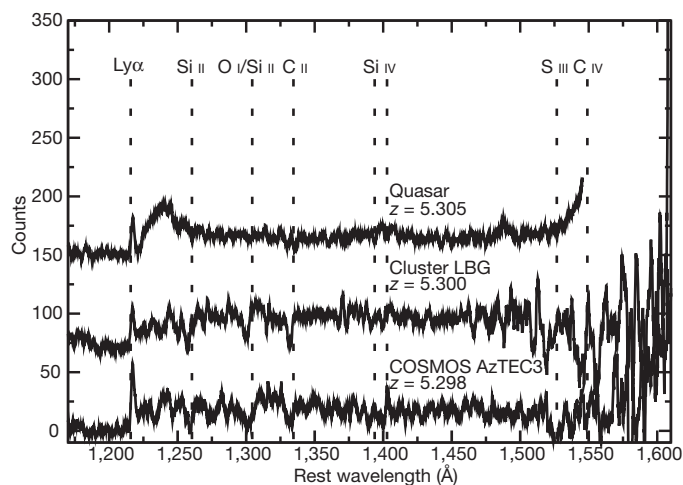


Figure 1 | Spectra of confirmed cluster members. These spectra were taken with the Keck II telescope and correspond to the extreme starburst (COSMOS AzTEC3), a combined spectrum of two Lyman-break galaxies at 95 kpc (Cluster LBG) and the Chandra-detected quasar at 13 Mpc from the extreme starburst. The galaxy spectra show absorption features indicative of interstellar gas (Si II, O I/Si II and C II) and young massive stars (Si IV and C IV) indicative of a stellar population less than 30 Myr old²⁶. The quasar shows broad Lyman- α ($\text{Ly}\alpha$) emission absorbed by strong winds, with a narrow Lyman- α line seen at the same systemic velocity as absorption features in the spectra.

¹Spitzer Science Centre, 314-6 California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA. ²Department of Astronomy, 249-17 California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA. ³National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA. ⁴Institut de Radio Astronomie Millimétrique, 300 rue de la Piscine, F-38406 St-Martin-d'Hères, France. ⁵Max-Planck-Institute für Plasma Physics, Boltzmann Strasse 2, Garching 85748, Germany. ⁶Max-Planck-Institute für Astronomie, Königstuhl 17, Heidelberg 69117, Germany. ⁷Department of Astronomy, University of Massachusetts, Lederle Graduate Research Tower B, 619E, 710 North Pleasant Street, Amherst, Massachusetts 01003-9305, USA. ⁸Harvard Smithsonian Center for Astrophysics, 60 Garden Street, MS. 67, Cambridge, Massachusetts 02138, USA. ⁹Department of Physics and Astronomy, University of California, Riverside, California 92521, USA. ¹⁰Johns Hopkins University, Laboratory for Observational Cosmology, Code 665, Building 34, NASA's Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

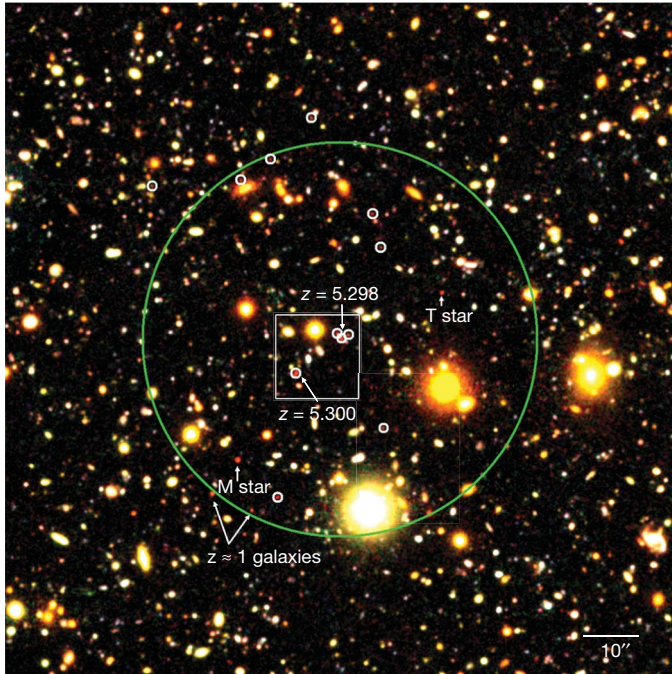


Figure 2 | Image of the region around the protocluster core. This area corresponds to a $2' \times 2'$ region around the starburst (COSMOS AzTEC-3). The $z \approx 5.3$ candidates are marked in white and a 2-Mpc co-moving radius is marked with a green circle. The boxed area is shown larger in Fig. 3, where the optical counterpart of the submillimetre source COSMOS AzTEC-3 is labelled 'Starburst'. Spectroscopic redshifts and other red objects that have been identified as galactic stars or low-redshift galaxies by their spectral energy distribution are also labelled.

observed limit on the submillimetre spectral slope favours models with colder dust and, hence, lower luminosities.

The significance of the overdensity around the starburst is immediately apparent in Figs 2 and 3. In the 1-square-arcmin area ($2.3 \times 2.3 \text{ Mpc}^2$ at $z = 5.3$) around the starburst, we would expect to find 0.75 ± 0.04 bright ($z_{850} < 26$) galaxies with colours consistent with a Lyman break in their spectra at $z = 5.3$ (ref. 19), but instead we find eight. This is an 11-fold overdensity, assuming the redshift range $4.5 < z < 6.5$ probed by typical broadband colour selections^{19,20}.

Within a 2-Mpc radius of the starburst, we find 11 objects brighter than L_* whose intermediate-band colours²¹ are consistent with being at $z = 5.3$. This represents a >11 -fold overdensity in both the measured and the expected density of luminous galaxies. Estimates of the typical variance from clustering and cosmological simulations suggest that this is significant at the $>9\sigma$ level even if we only consider the spectroscopically confirmed systems. Of these 11 objects, three (including the optical counterpart of COSMOS AzTEC-3) are within proper distance of 12.2 kpc ($2''$) of COSMOS AzTEC-3, and two additional spectroscopically confirmed sources are found 95 kpc ($15.5''$) away.

X-ray-selected (0.5–10-keV band) $z > 5$ quasars are extremely rare²² owing to the high luminosities required for detection, yet one is found¹⁷ within 13 Mpc of the starburst at the same spectroscopic redshift as COSMOS AzTEC-3. The distance between these objects is comparable to the co-moving distance scale expected for protoclusters at $z \approx 5$ (refs 5, 7). The optical spectrum of the quasar has deep, blue-shifted gas absorption features indicative of strong winds driven by the energy dissipated from the rapid black-hole growth. The object has an X-ray luminosity of $1.9 \times 10^{11} L_\odot$ and a bolometric luminosity estimated from its spectral energy distribution of $\geq 8.3 \times 10^{11} L_\odot$ (H. Hao *et al.*, manuscript in preparation), implying a black-hole mass of $\geq 3 \times 10^7 M_\odot$ if it is accreting at the Eddington rate, with a more likely mass of $\sim 3 \times 10^8 M_\odot$ for the typical accretion rate of one-tenth the

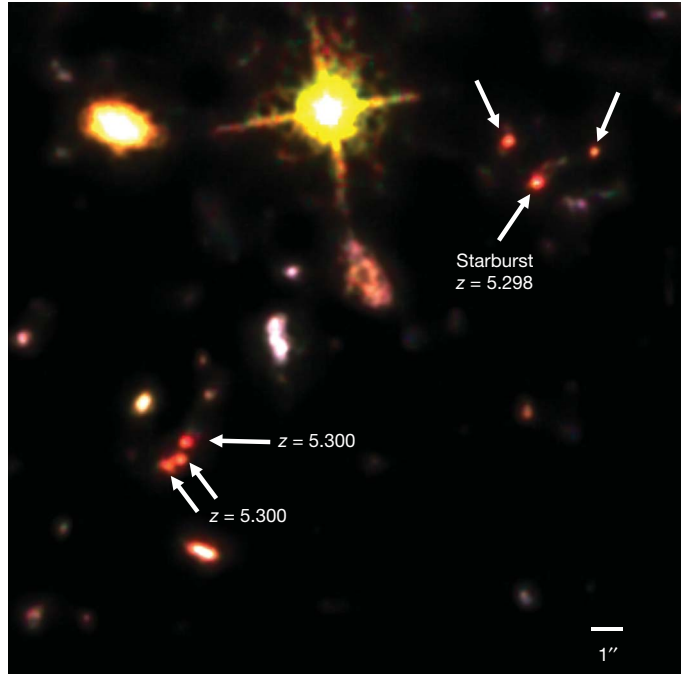


Figure 3 | Detail of the protocluster core. This area corresponds to a $22.5'' \times 22.5''$ area on the sky at a co-moving distance of 0.865 Mpc, or a proper distance of 0.137 Mpc, at $z = 5.298$. The six optically bright objects with spectral energy distributions consistent with $z = 5.298$ are marked and spectroscopic redshifts are indicated. The optical counterpart of the submillimetre source COSMOS AzTEC-3 is labelled 'Starburst'.

Eddington rate⁵. Assuming the final black hole/stellar mass relation to be $M_{\text{BH}} \approx 0.002 M_{\text{stars}}$ (ref. 5) implies that this object will eventually have a stellar mass of $>10^{10} M_\odot$ – $10^{11} M_\odot$, placing it among the most luminous and massive objects at this redshift^{19,23}.

We estimated the stellar mass of the protocluster system by fitting stellar population models to the rest-frame ultraviolet–optical photometry of the individual galaxies in the protocluster. The redshift was fixed at $z = 5.298$, and models with a single recent burst of star formation²⁴ were used, allowing for up to ten visual magnitudes of extinction²⁵. [O II] and H α emission lines were added to the templates with fluxes proportional to the ultraviolet continuum of the template¹⁸. The accuracy of the stellar mass estimate is limited by the sensitivity of the 0.9–2.5- μm photometry. The present data are insufficient to fully break the degeneracy between stellar age and dust obscuration. However, the age of 10 Myr derived from the photometric fitting is consistent with the features seen in the Keck spectra²⁶. Given the range of acceptable fits and the concordance with the Keck spectra, the resulting stellar mass is probably accurate to a factor of ~ 2 (0.3 dex).

Using the described procedure, we conservatively estimate that the starburst AzTEC-3 has a stellar mass of $(1\text{--}2) \times 10^{10} M_\odot$, implying that the baryonic matter is $>70\%$ gas, nearly twice that found in typical starburst systems²⁷ but in agreement with the dynamical estimates^{8,28}. The 11 objects in the protocluster core have a total stellar mass of $>2 \times 10^{10} M_\odot$, with individual galaxies weighing between $0.06 \times 10^9 M_\odot$ and $10 \times 10^9 M_\odot$. With this stellar mass and gas fraction, a lower limit can be placed on the total mass of this system, assuming a global dark matter/baryon ratio of 5.9 (ref. 1). The resulting total halo mass is $>4 \times 10^{11} M_\odot$, with the starburst residing in a halo of mass $>10^{11} M_\odot$, comparable to the halo masses predicted for galaxies that will eventually merge into present-day galaxy clusters⁷. However, we note that the actual mass is probably much higher because much of the baryonic mass is probably in unobserved hydrogen gas, and the starburst object alone accounts for $>37\%$ of the total mass. Furthermore, the contribution of significantly more numerous, fainter (luminosity, $< L_*$) galaxies¹⁹ are not counted in this mass estimate.

The three objects around COSMOS AzTEC-3 probably represent the progenitor of a massive central cluster galaxy (type cD) at lower redshift. These objects are already within the radius of a typical local cD galaxy and their dynamical timescale is ~ 60 Myr, assuming a velocity dispersion of 200 km s^{-1} . Even considering the objects at 95 kpc, the dynamical timescale is less than 0.5 Gyr, providing several dynamical times for a merger to occur by $z \approx 2$ (that is, 2 Gyr later). However, the observed stellar mass in these galaxies is significantly less than the $\sim 10^{11} M_{\odot}$ – $10^{12} M_{\odot}$ in a typical local cD galaxy⁷, indicating that the majority of the stars have yet to form.

The properties of this protocluster are in qualitative and quantitative agreement with galaxy formation simulations^{4,5}. The spatial extent, star formation rate per unit mass and gas properties of the core structure around the extreme starburst are all similar to the predictions for massive-galaxy formation in simulations. Furthermore, the properties of the quasar are also in agreement with the models of the later phases of massive-galaxy formation when the quasar becomes visible. Finally, unlike for previously described overdensities at $z > 5$ (ref. 6), we have strong spectroscopic and photometric evidence for a range of objects including massive, heavily star forming and active galaxies. These are found both in the core of the structure and over a much larger area, indicating that the effects of environment on galaxy formation as early as $z \approx 5$ can be studied. We conclude that this region contains a large-scale baryonic overdensity in the very early Universe that will evolve into a high-mass cluster like those observed at lower redshifts.

Received 18 May; accepted 16 November 2010.

Published online 12 January 2011.

- Larson, D. *et al.* Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: power spectra and WMAP-derived parameters. *Astrophys. J. Suppl. Ser.* (in the press); preprint at (<http://arxiv.org/abs/1001.4635>) (2010).
- Papovich, C. *et al.* A Spitzer-selected galaxy cluster at $z = 1.62$. *Astrophys. J.* **716**, 1503–1513 (2010).
- Mei, S. *et al.* Evolution of the color-magnitude relation in galaxy clusters at $z \sim 1$ from the ACS Intermediate Redshift Cluster Survey. *Astrophys. J.* **690**, 42–68 (2009).
- Springel, V. *et al.* Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636 (2005).
- Li, Y. *et al.* Formation of $z \sim 6$ quasars from hierarchical galaxy mergers. *Astrophys. J.* **665**, 187–208 (2007).
- Overzier, R. *et al.* Stellar masses of Lyman break galaxies, Ly α emitters, and radio galaxies in overdense regions at $z = 4$ – 6 . *Astrophys. J.* **704**, 548–563 (2009).
- Overzier, R. *et al.* Λ CDM predictions for galaxy protoclusters - I. The relation between galaxies, protoclusters and quasars at $z \sim 6$. *Mon. Not. R. Astron. Soc.* **394**, 577–594 (2009).
- Riechers, D. *et al.* A massive molecular gas reservoir in the $z = 5.3$ submillimeter galaxy AzTEC-3. *Astrophys. J. Lett.* **720**, 131–136 (2010).
- Robertson, B. *et al.* Photometric properties of the most massive high-redshift galaxies. *Astrophys. J.* **667**, 60–78 (2007).
- Miley, G. K. *et al.* A large population of ‘Lyman-break’ galaxies in a protocluster at redshift $z \approx 4.1$. *Nature* **427**, 47–50 (2004).
- Walter, F. *et al.* Molecular gas in the host galaxy of a quasar at redshift $z = 6.42$. *Nature* **424**, 406–408 (2003).
- Wang, R. *et al.* Molecular gas in $z \approx 6$ quasar host galaxies. *Astrophys. J.* **714**, 699–712 (2010).
- Scoville, N. Z. *et al.* The Cosmic Evolution Survey (COSMOS): overview. *Astrophys. J. Suppl. Ser.* **172**, 1–8 (2007).
- Scott, K. S. *et al.* AzTEC millimetre survey of the COSMOS field - I. Data reduction and source catalogue. *Mon. Not. R. Astron. Soc.* **385**, 2225–2238 (2008).
- Younger, J. D. *et al.* Evidence for a population of high-redshift submillimeter galaxies from interferometric imaging. *Astrophys. J.* **671**, 1531–1537 (2007).
- Schinnerer, E. *et al.* The VLA-COSMOS Survey. II. Source catalog of the large project. *Astrophys. J. Suppl. Ser.* **172**, 46–69 (2007).
- Elvis, M. *et al.* The Chandra COSMOS Survey. I. Overview and point source catalog. *Astrophys. J.* **184**, 158–171 (2009).
- Kennicutt, J. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–232 (1998).
- Bouwens, R. J., Illingworth, G. D., Franx, M. & Ford, H. UV luminosity functions at $z \sim 4$, 5, and 6 from the Hubble Ultra Deep Field and other deep Hubble Space Telescope ACS fields: evolution and star formation history. *Astrophys. J.* **670**, 928–958 (2007).
- Hildebrandt, H. *et al.* CARS: the CFHTLS-Archive-Research Survey. II. Weighing dark matter halos of Lyman-break galaxies at $z = 3$ – 5 . *Astron. Astrophys.* **498**, 725–736 (2009).
- Capak, P. *et al.* The first release COSMOS optical and near-IR data and catalog. *Astrophys. J. Suppl. Ser.* **172**, 99–116 (2007).
- Brusa, M. *et al.* High-redshift quasars in the COSMOS survey: the space density of $z > 3$ X-ray selected QSOs. *Astrophys. J.* **693**, 8–22 (2009).
- Stark, D. P. *et al.* The evolutionary history of Lyman break galaxies between redshift 4 and 6: observing successive generations of massive galaxies in formation. *Astrophys. J.* **697**, 1493–1511 (2009).
- Maraston, C. Evolutionary synthesis of stellar populations: a modular tool. *Mon. Not. R. Astron. Soc.* **300**, 872–892 (1998).
- Calzetti, D., Kinney, A. L. & Storchi-Bergmann, T. Dust obscuration in starburst galaxies from near-infrared spectroscopy. *Astrophys. J.* **458**, 132–135 (1996).
- Pettini, M., Steidel, C. C., Adelberger, K. L., Dickinson, M. & Giavalisco, M. The ultraviolet spectrum of MS 1512–CB58: an insight into Lyman-break galaxies. *Astrophys. J.* **528**, 96–107 (2000).
- Tacconi, L. *et al.* High molecular gas fractions in normal massive star-forming galaxies in the young Universe. *Nature* **463**, 781–784 (2010).
- Tacconi, L. J. *et al.* Submillimeter galaxies at $z \sim 2$: evidence for major mergers and constraints on lifetimes, IMF, and CO-H₂ conversion factor. *Astrophys. J.* **680**, 246–262 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements These results are based on observations with: the W. M. Keck Observatory, the IRAM Plateau de Bure Interferometer, the IRAM 30-m telescope with the GISMO 2-mm camera, the Chandra X-ray Observatory, the Subaru Telescope, the Hubble Space Telescope, the Canada-France-Hawaii Telescope with WIRC and MegaPrime, the United Kingdom Infrared Telescope, the Spitzer Space Telescope, the Smithsonian Submillimeter Array Telescope, the James Clerk Maxwell Telescope with the AzTEC 1.1mm camera, and the National Radio Astronomy Observatory's Very Large Array. D.R. and B.R. acknowledge support from NASA through Hubble Fellowship grants awarded by the Space Telescope Science Institute. P.L.C. and N.Z.S. acknowledge grant support from NASA. G.W.W., M.Y. and J.G.S. acknowledge grant support from the NSF.

Author Contributions P.L.C. led the spectroscopic effort, reduced the spectroscopic and photometric data, and led the scientific analysis including the optical and radio/millimetre fitting analysis and cluster properties. N.Z.S. led the spectroscopic and photometric follow-up efforts. D.R., C.C., P.C. and R.N. assisted with the physical interpretation of the radio data. B.R. provided cosmological simulations to check the significance of the protocluster and the likelihood of finding it. M.S., L.Y., M.E., F.C. and B.M. carried out the Keck observations and assisted with the data reduction. E.S. reduced and analysed the radio data. G.W.W. and M.Y. assisted with the submillimetre data analysis. F.C. and M.E. assisted with the X-ray data analysis. A.K. coordinated the 2-mm observations. J.G.S. conducted and reduced the 2-mm observations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.L.C. (capak@astro.caltech.edu).

Broadband waveguide quantum memory for entangled photons

Erhan Saglamyurek¹, Neil Sinclair¹, Jeongwan Jin¹, Joshua A. Slater¹, Daniel Oblak¹, Félix Bussi eres^{1†}, Mathew George², Raimund Ricken², Wolfgang Sohler² & Wolfgang Tittel¹

The reversible transfer of quantum states of light into and out of matter constitutes an important building block for future applications of quantum communication: it will allow the synchronization of quantum information¹, and the construction of quantum repeaters² and quantum networks³. Much effort has been devoted to the development of such quantum memories¹, the key property of which is the preservation of entanglement during storage. Here we report the reversible transfer of photon–photon entanglement into entanglement between a photon and a collective atomic excitation in a solid-state device. Towards this end, we employ a thulium-doped lithium niobate waveguide in conjunction with a photon-echo quantum memory protocol⁴, and increase the spectral acceptance from the current maximum⁵ of 100 megahertz to 5 gigahertz. We assess the entanglement-preserving nature of our storage device through Bell inequality violations⁶ and by comparing the amount of entanglement contained in the detected photon pairs before and after the reversible transfer. These measurements show, within statistical error, a perfect mapping process. Our broadband quantum memory complements the family of robust, integrated lithium niobate devices⁷. It simplifies frequency-matching of light with matter interfaces in advanced applications of quantum communication, bringing fully quantum-enabled networks a step closer.

Quantum communication is founded on the encoding of information, generally referred to as quantum information, into quantum states of light⁶. The resulting applications of quantum physics at its fundamental level offer cryptographic security through quantum key distribution without relying on unproved mathematical assumptions⁸ and allow for the disembodied transfer of quantum states between distant places by means of quantum teleportation⁶. Reversible mapping of quantum states between light and matter is central to advanced applications of quantum communication such as quantum repeaters² and quantum networks³, in which matter constitutes nodes that hold quantum information until needed, and thereby synchronize the information flow through the communication channel or network. Furthermore, such a quantum interface allows the generation of light–matter entanglement through the mapping of one of two entangled photons into matter. To determine whether and how different physical systems can be entangled, and to localize the fundamental or technological boundaries where this fascinating quantum link breaks down, are central goals in quantum physics and have received much attention over the past decades⁶.

The reversible light–matter interface can be realized through the direct transfer of quantum states from light onto matter and back, or through the generation of light–matter entanglement followed by teleportation of quantum information from an externally provided photon into matter, and eventually back. Experimental capabilities have advanced rapidly over the past years and quantum state transfer between light and atomic vapour^{9–13}, solid-state ensembles^{4,14}, or single absorbers¹⁵, as well as the generation of light–matter entanglement

through the absorption of photons^{16–18}, or the emission of photons from atomic ensembles^{19–21} or single emitters^{22,23} have all been reported.

For quantum memory to become practical, it is important to reduce the complexity of experimental implementations, and the recent addition of rare-earth-ion-doped crystals^{4,14} to the set of storage materials has been a valuable step towards this goal. The promise of such crystals is further enhanced through potentially long storage times—up to several seconds in Pr:Y₂SiO₅ (ref. 24). In addition, given the large inhomogeneous broadening of optical zero-phonon lines, up to ~100-gigahertz (GHz), rare-earth-ion-doped crystals in principle offer storage of photons with less than 100-picosecond duration when being used in conjunction with a suitable quantum memory protocol⁴. Yet, the reversible state transfer between light and solid-state devices has so far not been shown to preserve entanglement. This is largely due to the limited spectral bandwidth of current implementations, 100 megahertz (MHz) at most⁵, which is orders of magnitude smaller than that of entangled photon pairs generated in the widely used process of spontaneous parametric down-conversion⁶. In this work, we approach the problem from both ends: we increase the acceptance bandwidth of our storage device to 5 GHz and narrow the bandwidths of our entangled photons to similar values. Furthermore, by using a wave-guiding storage medium, we move fundamental quantum memory research further towards application.

The layout of our experiment is depicted in Fig. 1. Short pulses of 523-nm wavelength light travel through an unbalanced interferometer. For sufficiently small pulse energies, subsequent spontaneous parametric down-conversion yields, to a good approximation, individual pairs of photons, centred at wavelengths around 795 nm and 1,532 nm, in the time-bin entangled qubit state²⁵:

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}(|e,e\rangle + |l,l\rangle) \quad (1)$$

Here, $|e\rangle$ and $|l\rangle$ denote early and late temporal modes and replace the usual spin-down and spin-up notation for spin-half particles. More specifically, $|i,j\rangle$ denotes a quantum state in which the 795-nm photon has been created in the temporal mode i , and the 1,532-nm photon has been created in the temporal mode j . We point out that, owing to the spectral filtering, our source generates frequency-uncorrelated entangled photons at wavelengths that match the low-loss windows of free-space and standard telecommunication fibre. It can thus be readily used in real-world applications of quantum communication that involve quantum teleportation and entanglement swapping.

The 1,532-nm photon is directed to a qubit analyser. It consists of either a fibre delay line followed by a single-photon detector that monitors the photon's arrival time, or a fibre-optical interferometer that is unbalanced in the same way as the pump interferometer, followed by single-photon detectors. The role of the delay line is to perform projection measurements of the photon's state onto early and late qubit states. Alternatively, the interferometer enables projections onto

¹Institute for Quantum Information Science, and Department of Physics and Astronomy, University of Calgary, 2500 University Drive NW, Calgary, Alberta. T2N 1N4, Canada. ²Department of Physics—Applied Physics, University of Paderborn, Warburger Strasse 100, 33095 Paderborn, Germany. [†]Present address: GAP-Optique, University of Geneva, Rue de l' cole-de-M decine 20, 1211 Geneva 4, Switzerland.

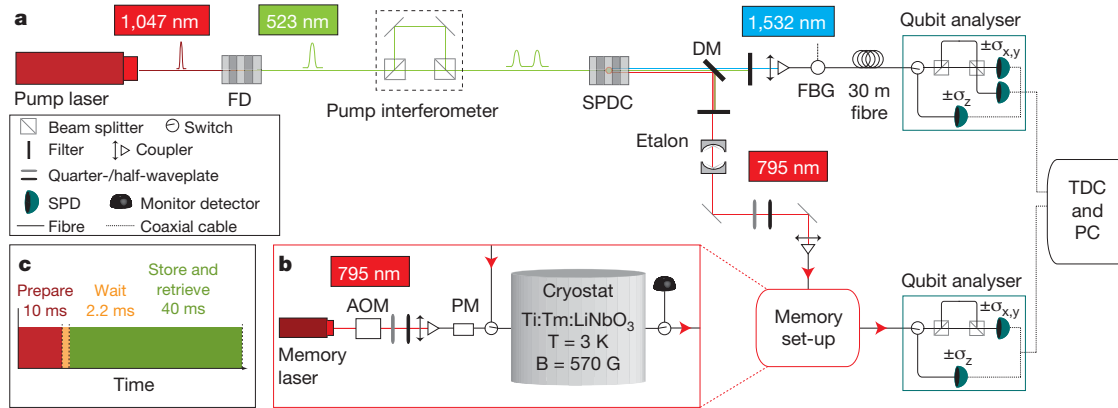


Figure 1 | Schematics of the experimental set-up. **a**, Generating and measuring entanglement. Six-picosecond-long pump laser pulses (1,047.328 nm wavelength, 80 MHz repetition rate) are frequency doubled (FD) in a periodically poled lithium niobate (PPLN) crystal. Each resulting 16-ps-long pulse (523.664 nm wavelength, 90 mW average power) is coherently split into two by the unbalanced pump interferometer, featuring a 1.4-ns travel-time difference. Spontaneous parametric down-conversion (SPDC) in a second PPLN crystal followed by frequency filtering using an etalon and a fibre Bragg grating (FBG) (bandwidths of 6 GHz and 9 GHz, respectively), yields maximally entangled pairs of photons centred at 795.506 nm and 1,532.426 nm wavelength (DM, dichroic mirror). The 1,532-nm photon travels through a 30-m telecommunication fibre, and the 795-nm photon is either stored in the memory or sent through a fibre delay line (not pictured). To characterize the bi-photon state, we use qubit analysers consisting of delay lines or unbalanced interferometers connected to single-photon detectors. Detection events are

equal superpositions of early and late modes²⁵. Using the language of spin-half systems, this corresponds to projections onto σ_z and, for appropriately chosen phases, σ_x and σ_y , respectively.

The 795-nm photon is transmitted to the quantum memory where its state—specifically that it is entangled with the 1,532-nm photon—is mapped onto a collective excitation of millions of thulium ions. Some time later, the state is mapped back onto a photon that exits the memory through a fibre in well-defined spatio-temporal modes and is probed by a second qubit analyser.

collected with a time-to-digital converter (TDC) connected to a personal computer (PC). All interferometers are phase-locked to stable reference lasers (not shown). **b**, Memory set-up. The 795.506-nm continuous-wave memory laser beam is intensity- and phase/frequency-modulated using an acousto-optic modulator (AOM) and a phase modulator (PM). The waveguide is cooled to 3 K and exposed to a 570-G magnetic field aligned with the crystal's C_3 -axis. Waveplates allow adjusting the polarization of the beam to the waveguide's transverse magnetic (TM) mode, and optical switches combine and separate the optical pump beam and the 795-nm photons. **c**, Timing sequence. We use three continuously repeated phases: the 10 ms 'prepare' phase for optical pumping, the 2.2-ms 'wait' phase, which ensures stored photons are not polluted by fluorescence from the excited state, and the 40-ms 'store and retrieve' phase, during which many 795-nm photons are successively stored in the waveguide and recalled after 7 ns.

To reversibly map the 795-nm photon onto matter, we use a photon-echo quantum memory protocol based on atomic frequency combs (AFC)⁴. It is rooted in the interaction of light with an ensemble of atomic absorbers (so far rare-earth-ion-doped crystals cooled to cryogenic temperatures) with an inhomogeneously broadened absorption line that has been tailored into a series of equally spaced absorption peaks (see Fig. 2). The absorption of a single photon leads to a collective excitation shared by many atoms. Owing to the particular shape of the tailored absorption line, the excited collective coherence rapidly dephases and repeatedly recovers after multiples of the storage time T_s . This results in the re-emission of a photon in the state encoded into the original photon.

In our implementation the moment of photon re-emission is predetermined by the spacing of the teeth in the comb, $T_s = 1/\Delta$, and the storage process can be described as arising from the linear response of an optical filter made by spectral hole burning. Yet, readout on demand can be achieved by temporarily mapping the optically excited coherence onto ground-state coherence where the comb spacing is smaller or the comb structure is washed out⁴, or by combining the AFC protocol with controlled reversible inhomogeneous broadening of each absorption line, similar to the storage mechanism used in another photon-echo quantum memory protocol¹.

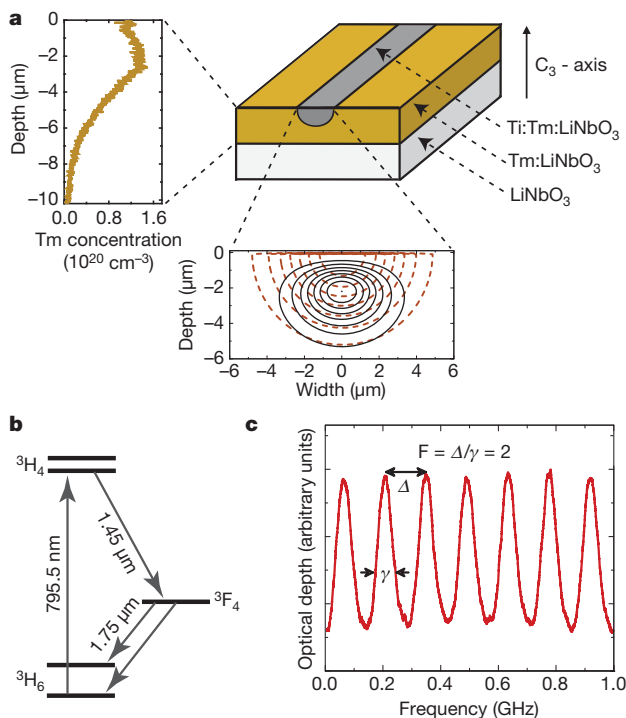


Figure 2 | The storage medium. **a**, Waveguide geometry. The measured thulium (Tm) concentration profile is given on the left and the calculated intensity distribution of the fundamental TM-mode at the 795-nm wavelength is shown below. Iso-intensity lines are plotted corresponding to 100%, 87.5%, 75% and so on of the maximum intensity. **b**, Simplified energy level diagram of thulium ions. The optical coherence time of the $^3H_6 \leftrightarrow ^3H_4$ transition at 3 K is 1.6 μs, the radiative lifetimes of the 3H_4 and 3F_4 levels are 82 μs and 2.4 ms, respectively, and the branching ratio from the 3H_4 to the 3F_4 level is 44%. Upon application of a magnetic field of 570 G, the ground and excited levels split into magnetic sublevels with lifetimes exceeding one second²⁷. **c**, Atomic frequency comb. The bandwidth of our AFC is 5 GHz (shown here is a 1-GHz broad section). The separation between the teeth is $\Delta \approx 143$ MHz, resulting in 7 ns storage time. The line width of the peaks is $\gamma \approx 75$ MHz, yielding a finesse $F = 2$, as expected for the sinus-type comb.

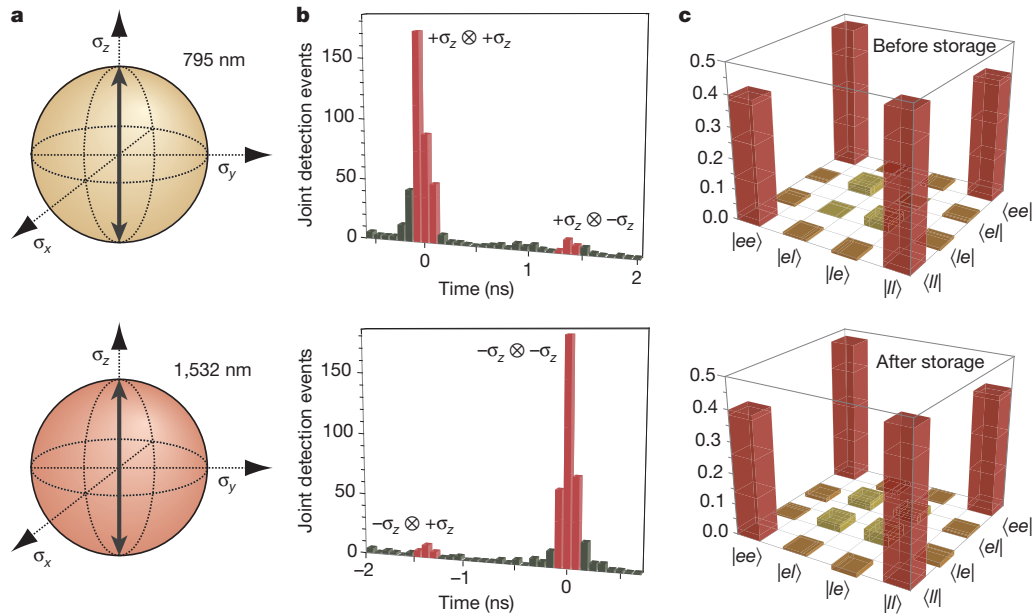


Figure 3 | Measurement of density matrices. **a**, Visualization of projection measurements. The measurement settings for the 795-nm (or 1,532-nm) qubit analyser are depicted on the upper (or lower) Bloch sphere. The example shows joint settings that enable calculating normalized probabilities for projections onto $\sigma_z \otimes \sigma_z$ and $\sigma_z \otimes -\sigma_z$. **b**, Results for joint projection measurement after storage. The top (bottom) histogram displays joint detection events for the projection onto $\sigma_z \otimes \sigma_z$ and $\sigma_z \otimes -\sigma_z$ ($-\sigma_z \otimes \sigma_z$ and $-\sigma_z \otimes -\sigma_z$) as a function of the time difference between detections of the 795-nm and the 1,532-nm

photons. Our storage device, a Ti:Tm:LiNbO₃ optical waveguide cooled to 3 K, is detailed in Fig. 2. It was previously characterized to establish its suitability as a photon-echo quantum memory material²⁶. It combines interesting properties from the specific rare-earth element (795-nm storage wavelength), the host crystal (allowing for controlled dephasing and rephasing by means of electric fields), and from the waveguiding structure (ease-of-use). Lithium niobate waveguides have also been doped with neodymium, praseodymium and erbium⁷, and we conjecture that other rare-earth ions could also be used. This could extend the properties of LiNbO₃ and allow an integrated approach to other storage wavelengths, ions with different level structures, and so on.

To generate the AFC, we use a sideband-chirping technique (see Supplementary Information) to transfer atomic population between magnetic sublevels and create troughs and peaks in the inhomogeneously broadened absorption line. They form a 5-GHz-wide comb with tooth spacing of 143 MHz, setting the storage time to 7 ns. The system efficiency in our implementation is currently about 0.2%. This is in part due to the 90% fibre-to-waveguide input and output coupling loss, which we attribute to imperfect mode overlap. In addition, owing to the specific level structure of thulium under current experimental conditions, the finesse of the comb in the broadband approach is two, which limits the memory efficiency to about 10%. However, imperfections in the creation of the comb decrease this efficiency to around 2%. The system efficiency can be increased by improving the spectral tailoring of the AFC, and triggering photon re-emission in the backward direction. By also optimizing the mode overlap, we anticipate that it could reach

approximately 15%. Furthermore, if the two long-lived atomic levels between which population is transferred during the optical pumping procedure (in our case the two magnetic ground states; see Fig. 2) are spaced by more than the storage bandwidth, the memory efficiency can theoretically reach unity (see Supplementary Information).

To assess the quantum nature of our light-matter interface, we first make projection measurements with the 795-nm photons and the 1,532-nm photons onto time-bin qubit states characterized by Bloch vectors aligned along **a** and **b**, respectively, where **a**, **b** ∈ [$\pm\sigma_x, \pm\sigma_y, \pm\sigma_z$] (see Fig. 3). Experimentally, this is done by means of suitably adjusted qubit analysers, and by counting the number $C(\mathbf{a}, \mathbf{b})$ of detected photon pairs. From two such spin-measurements, we calculate the normalized joint-detection probability:

$$P(\mathbf{a}, \mathbf{b}) = \frac{C(\mathbf{a}, \mathbf{b})}{C(\mathbf{a}, \mathbf{b}) + C(\mathbf{a}, -\mathbf{b})} \quad (2)$$

The measurement and the results with the fibre delay line, as well as the memory, are detailed in Fig. 3 and the Supplementary Information. From this data, we reconstruct the bi-photon states before and after storage in terms of their density matrices ρ_{in} and ρ_{out} , depicted in Fig. 3, using a maximum likelihood estimation²⁷. This, in turn, allows us to examine the entanglement of formation²⁸, a measure that indicates entanglement if it exceeds zero; it is upper-bounded by one. The results, listed in Table 1, clearly show the presence of entanglement in ρ_{in} and ρ_{out} and, within experimental uncertainty, establish that the storage process preserves entanglement without measurable degradation.

Table 1 | Entanglement measures, purities and fidelities

| | Entanglement of formation (%) | Purity (%) | Fidelity with $ \phi^+\rangle$ (%) | Input/output fidelity (%) | Expected S_{th} | Measured S |
|---------------------|-------------------------------|----------------|------------------------------------|---------------------------|--------------------------|-------------------|
| ρ_{in} | 64.4 ± 4.2 | 75.7 ± 2.4 | 86.2 ± 1.5 | | 2.235 ± 0.085 | 2.379 ± 0.034 |
| ρ_{out} | 65 ± 11 | 76.3 ± 5.9 | 86.6 ± 3.9 | 95.4 ± 2.9 | 2.2 ± 0.22 | 2.25 ± 0.06 |

Entanglement of formation (normalized with respect to the entanglement of formation of $|\phi^+\rangle$), purity $P = \text{tr}(\rho^2)$, fidelity with $|\phi^+\rangle$, input-output fidelity $F = \left(\text{tr} \sqrt{\sqrt{\rho_{\text{out}}} \rho_{\text{in}} \sqrt{\rho_{\text{out}}}} \right)^2$ (referring to the fidelity of ρ_{out} with respect to ρ_{in}), and expected and experimentally obtained S values for tests of the CHSH Bell inequality (measured for $\mathbf{a} = \sigma_x$, $\mathbf{a}' = \sigma_y$, $\mathbf{b} = \sigma_x + \sigma_y$, and $\mathbf{b}' = \sigma_x - \sigma_y$). The correlation coefficients used to compute S and the calculation of S_{th} are detailed in the Supplementary Information. We note that the original state (and hence the recalled state) has limited purity and fidelity with $|\phi^+\rangle$. This is due to the probabilistic nature of our spontaneous parametric down-conversion source, which features a non-negligible probability of generating more than two photons simultaneously²⁶. Uncertainties indicate one-sigma standard deviations and are estimated from Poissonian detection statistics and using a Monte Carlo simulation.

Furthermore, we note that the fidelity F between ρ_{in} and ρ_{out} is close to one, and hence the unitary transformation introduced by the storage process is almost the identity transformation.

In addition, as a second entanglement measure, we perform tests of the Clauser–Horne–Shimony–Holt (CHSH) Bell inequality⁶. This test indicates non-local correlations and thus the possibility of using the bi-photons for entanglement-based quantum key distribution⁸ if the sum:

$$S = |E(\mathbf{a}, \mathbf{b}) + E(\mathbf{a}', \mathbf{b}) + E(\mathbf{a}, \mathbf{b}') - E(\mathbf{a}', \mathbf{b}')| \quad (3)$$

of four correlation coefficients

$$E(\mathbf{a}, \mathbf{b}) = \frac{C(\mathbf{a}, \mathbf{b}) - C(\mathbf{a}, -\mathbf{b}) - C(-\mathbf{a}, \mathbf{b}) + C(-\mathbf{a}, -\mathbf{b})}{C(\mathbf{a}, \mathbf{b}) + C(\mathbf{a}, -\mathbf{b}) + C(-\mathbf{a}, \mathbf{b}) + C(-\mathbf{a}, -\mathbf{b})} \quad (4)$$

with appropriately chosen settings \mathbf{a}, \mathbf{a}' and \mathbf{b}, \mathbf{b}' exceeds the classical bound of two; quantum mechanically it is upper-bounded by $2\sqrt{2}$. As detailed in Table 1, we find $S_{\text{in}} = 2.379 \pm 0.034 > 2$ before the memory and, crucially, $S_{\text{out}} = 2.25 \pm 0.06 > 2$, which is in agreement with the value $S_{\text{th}} = 2.2 \pm 0.22$ predicted from the reconstructed density matrix ρ_{out} . This validates the suitability of our set-up for quantum communication.

Our investigation provides an example of entanglement being transferred between physical systems of different nature, thereby adding evidence that this fundamental quantum property is not as fragile as is often believed. Furthermore, our broadband integrated approach permits the linkage of a promising quantum storage device with extensively used, high-performance sources of photons in bi- and multi-partite entangled states⁶. Although the storage efficiency and the storage time need to be significantly increased, and the moment of recall was pre-set, this study opens the way to new investigations of fundamental and applied aspects of quantum physics. Having increased the storage bandwidth also significantly facilitates the building of future quantum networks, because mutual frequency matching of photons and distant quantum memories will be simple. In addition, a large storage bandwidth—that is, the possibility to encode quantum information into short optical pulses—allows us to increase the number of temporal modes that can be stored during a given time. This enhances the flow of quantum information through a network and decreases the time needed to establish entanglement over a large distance using a quantum repeater^{1,2}.

We note that, parallel to this work, Clausen *et al.* have demonstrated the storage of an entangled photon using a neodymium-doped crystal²⁹.

Received 1 September; accepted 6 December 2010.

Published online 12 January 2011.

1. Lvovsky, A. I., Sanders, B. C., & Tittel, W. Optical quantum memory. *Nature Photon.* **3**, 706–714 (2009).
2. Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. Preprint at (<http://arxiv.org/abs/0906.2699>) (2009).
3. Kimble, H. J. The quantum Internet. *Nature* **453**, 1023–1030 (2008).
4. de Riedmatten, H., Afzelius, M., Staudt, M. U., Simon, C. & Gisin, N. A solid-state light–matter interface at the single-photon level. *Nature* **456**, 773–777 (2008).
5. Usmani, I., Afzelius, M., de Riedmatten, H. & Gisin, N. Mapping multiple photonic qubits into and out of one solid-state atomic ensemble. *Nature Commun.* **1**, 1–7 (2010).
6. Pan, J.-W., Chen, Z.-B., Żukowski, M., Weinfurter, H. & Zeilinger, A. Multi-photon entanglement and interferometry. Preprint at (<http://arxiv.org/abs/0805.2853>) (2008).

7. Sohler, W. *et al.* Integrated optical devices in lithium niobate. *Opt. Photon. News* 24–31 (January 2008).
8. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
9. Julsgaard, B., Sherson, J. & Cirac, J. I. J. Fűrřšek, J. & Polzik, E. S. Experimental demonstration of quantum memory for light. *Nature* **432**, 482–486 (2004).
10. Chanelière, T. *et al.* Storage and retrieval of single photons transmitted between remote quantum memories. *Nature* **438**, 833–836 (2005).
11. Eisaman, M. D. *et al.* Electromagnetically induced transparency with tunable single-photon pulses. *Nature* **438**, 837–841 (2005).
12. Honda, K. *et al.* Storage and retrieval of a squeezed vacuum. *Phys. Rev. Lett.* **100**, 093601 (2008).
13. Appel, J., Figueroa, E., Korystov, D., Lobino, M. & Lvovsky, A. Quantum memory for squeezed light. *Phys. Rev. Lett.* **100**, 093602 (2008).
14. Hedges, M. P., Longdell, J. J., Li, Y. & Sellars, M. J. Efficient quantum memory for light. *Nature* **465**, 1052–1056 (2010).
15. Boozer, A. D. *et al.* Reversible state transfer between light and a single trapped atom. *Phys. Rev. Lett.* **98**, 193601 (2007).
16. Choi, C. S. & Deng, H. Laurat, J. & Kimble, H. J. Mapping photonic entanglement into and out of a quantum memory. *Nature* **452**, 67–71 (2008).
17. Akiba, K. & Kashiwagi, K. Arikawa, M. & Kozuma, M. Storage and retrieval of non-classical photon pairs and conditional single photons generated by the parametric down-conversion process. *N. J. Phys.* **11**, 013049 (2009).
18. Jin, X.-M. *et al.* Quantum interface between frequency-uncorrelated down-converted entanglement and atomic-ensemble quantum memory. Preprint at (<http://arxiv.org/abs/1004.4691>) (2010).
19. Chou, C. W. *et al.* Measurement-induced entanglement for excitation stored in remote atomic ensembles. *Nature* **438**, 828–832 (2005).
20. Matsukevich, D. N. *et al.* Entanglement of a photon and a collective atomic excitation. *Phys. Rev. Lett.* **95**, 040405 (2005).
21. Yuan, Z.-S. *et al.* Experimental demonstration of a BDCZ quantum repeater node. *Nature* **454**, 1098–1101 (2008).
22. Blinov, B. B., Moehring, D. L., Duan, L.-M. & Monroe, C. Observation of entanglement between a single trapped atom and a single photon. *Nature* **428**, 153–157 (2004).
23. Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
24. Longdell, J., Fraval, E., Sellars, M. & Manson, N. Stopped light with storage times greater than one second using electromagnetically induced transparency in a solid. *Phys. Rev. Lett.* **95**, 063601 (2005).
25. Marcikic, I. *et al.* Time-bin entangled qubits for quantum communication created by femtosecond pulses. *Phys. Rev. A* **66**, 062308 (2002).
26. Sinclair, N. *et al.* Spectroscopic investigations of a Ti:TM:LiNbO₃ waveguide for photon-echo quantum memory. *J. Lumin.* **130**, 1586–1593 (2010).
27. Altepeter, J. B., Jeffrey, E. R., & Kwiat, P. G. Photonic state tomography. *Adv. At. Mol. Opt. Phys.* **52**, 105–159 (2005).
28. Plenio, M. B. & Virmani, S. An introduction to entanglement measures. *Quant. Inf. Comput.* **7**, 1–51 (2007).
29. Clausen, C. *et al.* Quantum storage of photonic entanglement in a crystal. *Nature* doi:10.1038/nature09662 (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is supported by NSERC, QuantumWorks, General Dynamics Canada, iCORE (now part of Alberta Innovates), CFI, AAET and FQRNT. We thank C. La Mela, T. Chanelière, T. Stuart, V. Kiselyov and C. Dascolas for help during various stages of the experiment, C. Simon, K. Rupavatham and N. Gisin for discussions, and A. Lvovsky for lending us a single-photon detector.

Author Contributions The Ti:TM:LiNbO₃ waveguide was fabricated and characterized at room temperature by M.G., R.R. and W.S. The photon-pair source was built by J.J., J.A.S. and F.B., the AFC memory set-up was developed by E.S. and N.S., and the complete experiment was conceived and directed by W.T. The measurements and the analysis were done by E.S., N.S., J.J., J.A.S., D.O. and W.T., and W.T., E.S., N.S., J.J., J.A.S. and D.O. wrote the paper. E.S., N.S., J.J. and J.A.S. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.T. (wtittel@ucalgary.ca).

An actively accreting massive black hole in the dwarf starburst galaxy Henize 2-10

Amy E. Reines¹, Gregory R. Sivakoff¹, Kelsey E. Johnson^{1,2} & Crystal L. Brogan²

Supermassive black holes are now thought to lie at the heart of every giant galaxy with a spheroidal component, including our own Milky Way^{1,2}. The birth and growth of the first ‘seed’ black holes in the earlier Universe, however, is observationally unconstrained³ and we are only beginning to piece together a scenario for their subsequent evolution⁴. Here we report that the nearby dwarf starburst galaxy Henize 2-10 (refs 5 and 6) contains a compact radio source at the dynamical centre of the galaxy that is spatially coincident with a hard X-ray source. From these observations, we conclude that Henize 2-10 harbours an actively accreting central black hole with a mass of approximately one million solar masses. This nearby dwarf galaxy, simultaneously hosting a massive black hole and an extreme burst of star formation, is analogous in many ways to galaxies in the infant Universe during the early stages of black-hole growth and galaxy mass assembly. Our results confirm that nearby star-forming dwarf galaxies can indeed form massive black holes, and that by implication so can their primordial counterparts. Moreover, the lack of a substantial spheroidal component in Henize 2-10 indicates that supermassive black-hole growth may precede the build-up of galaxy spheroids.

The starburst in Henize 2-10, a relatively nearby (9 megaparsecs, ~ 30 million light years) blue compact dwarf galaxy, has attracted the attention of astronomers for decades^{6–10}. Stars are forming in Henize 2-10 at a prodigious rate^{8,11,12} that is ten times that of the Large Magellanic Cloud¹³ (a satellite galaxy of the Milky Way), despite the fact that both of these dwarf galaxies have similar stellar masses^{14–16} and neutral hydrogen gas masses^{7,17}. Most of the star formation in Henize 2-10 is concentrated in a large population of very massive and dense ‘super-star clusters’, the youngest having ages of a few million years and masses of one hundred thousand times the mass of the Sun⁶. The main optical body of the galaxy has an extent less than a kiloparsec ($\sim 3,000$ light-years) in size and has a compact irregular morphology typical of blue compact dwarfs (Fig. 1).

We observed Henize 2-10 at centimetre radio wavelengths with the Very Large Array (VLA) and in the near-infrared with the Hubble Space Telescope (HST) as part of a large-scale panchromatic study of nearby dwarf starburst galaxies harbouring infant super-star clusters^{18–20}. A comparison between the VLA and HST observations drew our attention to a compact ($< 24 \text{ pc} \times 9 \text{ pc}$) central radio source located between two bright regions of ionized gas (Fig. 2). These data exclude any association of this central radio source with a visible stellar cluster (Fig. 3; see Supplementary Information for a discussion of the astrometry). Furthermore, the radio emission from this source has a significant non-thermal component ($\alpha \approx -0.4$, $S_\nu \propto \nu^\alpha$ where S_ν is the flux density at frequency ν) between 4.9 GHz and 8.5 GHz, as noted in previous studies of the galaxy⁹. An archival observation of Henize 2-10 taken with the Chandra X-ray Observatory reveals that a point source with hard X-ray emission is also coincident (to within the position uncertainty) with the central non-thermal radio source¹⁰ (see Supplementary Information). Typically, even powerful non-nuclear radio and X-ray sources (for example, supernova remnants and active X-ray binaries)

are at least an order of magnitude less luminous than the central source in Henize 2-10 (see Supplementary Information). In contrast, the radio and hard X-ray luminosities of the central source in Henize 2-10, as well as their ratio, are similar to known low-luminosity active galactic nuclei powered by accretion onto a massive black hole²¹.

The central, compact, non-thermal radio source in Henize 2-10 is also coincident with a local peak in Pa α and H α emission and appears to be connected to a thin quasi-linear ionized structure between two bright and extended regions of ionized gas. This morphology is tantalizingly suggestive of outflow (Fig. 2). Although we cannot conclusively determine whether or not this linear structure is physically connected to the brightest emitting regions with the data in hand, ground-based spectroscopic observations²² confirm a coherent velocity gradient along the entire ionized gas structure seen in Fig. 2, consistent with outflow or rotation. Moreover, a comparison between the central velocity of this ionized gas structure and the systemic velocity of the galaxy—derived

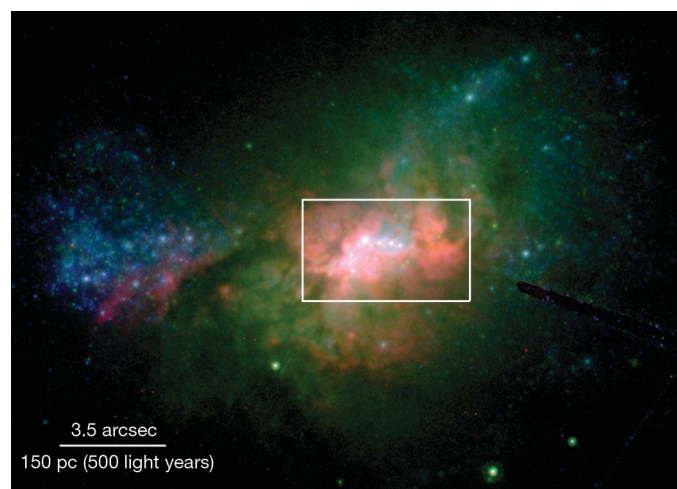


Figure 1 | Henize 2-10. Henize 2-10 is a blue compact dwarf galaxy hosting a concentrated region of extreme star formation. Using H α (ref. 8) and 24 μm (ref. 11) fluxes from the literature, we estimate a star-formation rate¹² of $1.9 M_\odot \text{ yr}^{-1}$, assuming that all of the emission is from the starburst and that the contribution from the active nucleus is negligible. We estimate that Henize 2-10 has a stellar mass of $3.7 \times 10^9 M_\odot$ from the integrated 2MASS K $_s$ -band flux^{14,15}. Neutral hydrogen observations of Henize 2-10 indicate a solid-body rotation curve typical of dwarf galaxies with a maximum projected rotational velocity of 39 km s^{-1} relative to the systemic velocity of the galaxy⁷. These observations also indicate a dynamical mass of about $10^{10} M_\odot$ within 2.1 kiloparsecs (ref. 7). The main optical body of the galaxy, shown here, is less than one kiloparsec across. Henize 2-10 shows signs of having undergone an interaction, including tidal-tail-like features in both its gaseous⁷ and stellar distributions (seen here). In this three-colour HST image of the galaxy, we show ionized gas emission in red (H α) and the stellar continuum in green ($\sim I$ -band, $0.8 \mu\text{m}$) and blue ($\sim U$ -band, $0.3 \mu\text{m}$). These archival data were taken with Wide Field and Planetary Camera 2 (H α) and the Advanced Camera for Surveys (U - and I -band). The white box indicates the region shown in Figs 2 and 3.

¹Department of Astronomy, University of Virginia, 530 McCormick Road, Charlottesville, Virginia 22904, USA. ²National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22904, USA.

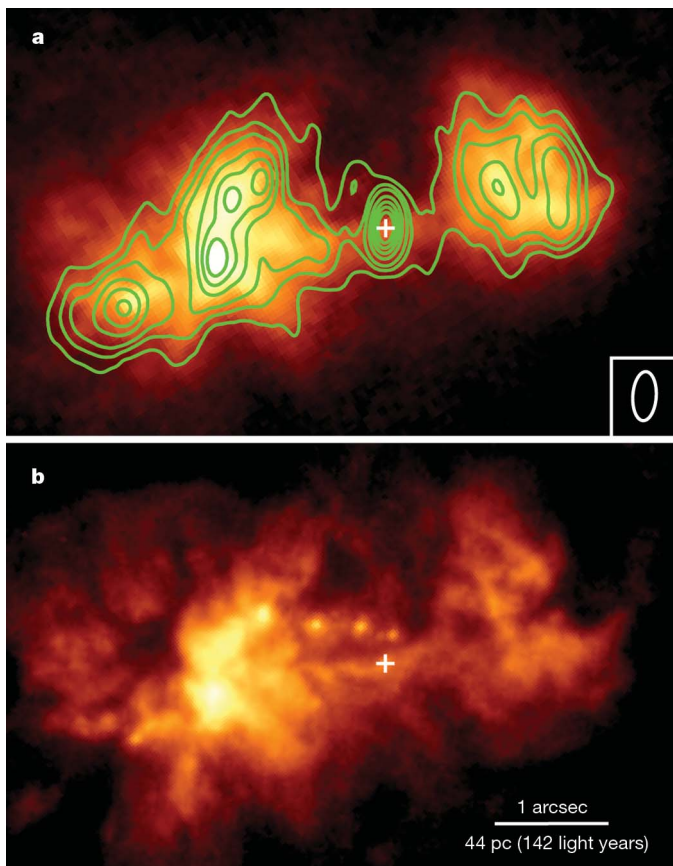


Figure 2 | The active nucleus and ionized gas in Henize 2-10. The overall morphology of the radio emission (green contours) in the central region of Henize 2-10 matches that of the ionized gas detected with HST (colour images), suggesting a shared origin. The active nucleus (plus symbol) is detected as a non-thermal nuclear VLA radio source coincident with a Chandra point source with hard X-ray emission. The nuclear source is also coincident with a local peak of ionized gas emission and appears to be connected to the thin quasi-linear structure between the two bright and extended regions of ionized gas: this is suggestive of, although not proof of, outflow. The central source has 4.9 GHz and 8.5 GHz radio luminosities of $7.4 \times 10^{35} \text{ erg s}^{-1}$ and $1.0 \times 10^{36} \text{ erg s}^{-1}$, respectively. The X-ray luminosity of the central source in the 2–10 keV band is $\sim 2.7 \times 10^{39} \text{ erg s}^{-1}$. The accretion rate of the $2 \times 10^6 M_{\odot}$ black hole is $\sim 5 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$ assuming an X-ray bolometric correction of 10 and an accretion efficiency of 0.1. **a**, Narrowband imaging with the Near Infrared Camera and Multi-Object Spectrometer (NICMOS) on the Hubble Space Telescope was used to trace the ionized gas in Henize 2-10 using the Pa α hydrogen recombination line at 1.87 μm . Continuum emission was removed using a neighbouring off-line narrowband filter. VLA 8.5 GHz (3.5 cm) radio contours are over-plotted in green and the active galactic nucleus is marked with a plus symbol. Contour levels are 9, 13, 17, 25, 33, 41 and 49 times the root-mean-square noise (12 μJy per beam). The beam is shown in the lower right corner. **b**, Optical narrowband imaging of the H α hydrogen recombination line at 0.66 μm yields a higher-resolution view of the ionized gas in Henize 2-10. The continuum has not been subtracted in this archival image, leaving young star clusters also visible.

from observations of neutral hydrogen gas rotating as a solid body⁷—indicates that the position of the non-thermal radio source is consistent with the dynamical centre of the galaxy.

Compact radio and hard X-ray emission at the centre of a galaxy are generally good indicators of accretion onto a massive black hole²¹, but we have also considered alternative explanations for the data. As discussed at length in the Supplementary Information, it is extremely unlikely that the nuclear source in Henize 2-10 is one or more supernova remnants, more recently created supernovae, stellar-mass black-hole X-ray binaries, or some combination of these phenomena. Briefly, X-ray binaries are too weak in the radio, supernova remnants are too weak in hard X-rays, and young compact radio supernovae are ruled out

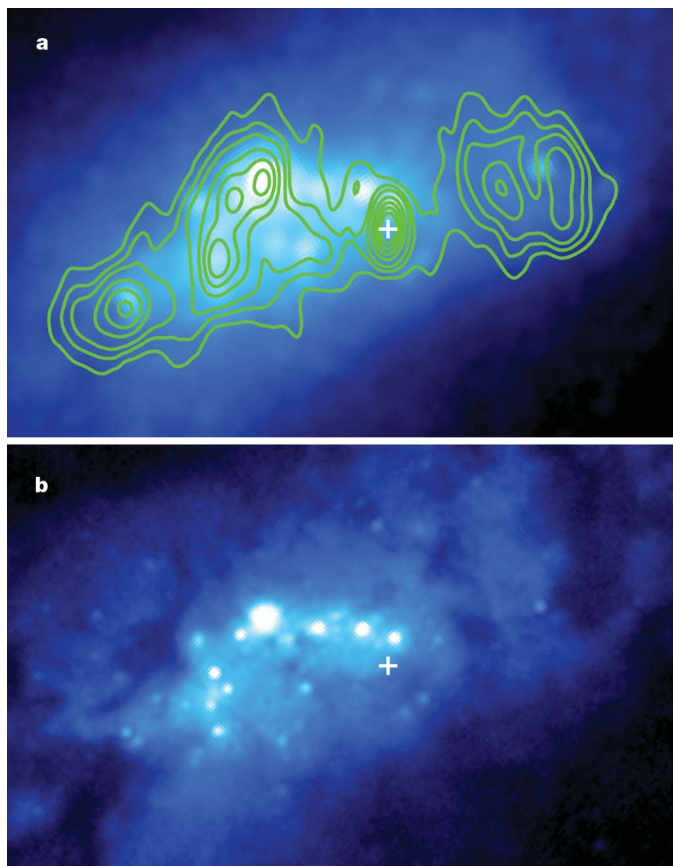


Figure 3 | Young super-star clusters in Henize 2-10. The overall structure of the radio emission (green contours) differs markedly from the distribution of star clusters in the centre of the galaxy (colour images). In particular, the non-thermal nuclear radio source does not have a detectable counterpart in these broadband continuum images (plus symbol), excluding any association with a visible star cluster. **a**, A near-infrared ($\sim K$ -band) image of the central region of Henize 2-10 overplotted with the same radio contours as in Fig. 2. HST/NICMOS was used to observe the galaxy through a broadband filter centred at 2.1 μm , which primarily traces the distribution of stellar light. **b**, A higher-resolution view of the star clusters is shown in this archival 0.8 μm ($\sim I$ -band) broadband image. The field of view is the same as in Fig. 2.

by observations using Very Long Baseline Interferometry²³. Although it may be possible to account for the radio and X-ray luminosities of the nuclear source with just the right combination of the abovementioned phenomena, the probability of such a coincidence is exceedingly low (see Supplementary Information). On the contrary, the radio and hard X-ray luminosities of the central source in Henize 2-10 are well within the range of known low-luminosity active galactic nuclei²¹.

In addition to ruling out young compact supernovae, the non-detection of the nuclear radio source at higher resolution ($\sim 0.5 \text{ pc} \times 0.1 \text{ pc}$) using Very Long Baseline Interferometry²³ may also seemingly rule out the presence of an actively accreting massive black hole. However, Seyfert nuclei with steep radio spectra ($\alpha \lesssim -0.5$, $S_{\nu} \propto \nu^{\alpha}$) often exhibit this ‘missing flux’ phenomenon when observed at increasingly higher spatial resolution²⁴. In these active galactic nuclei, as much as $\sim 90\%$ of the radio emission is absent on parsec scales and is expected to be dominated by extended low-surface-brightness features on larger scales, such as jets. This is in contrast to Seyferts with flat or positive radio spectra ($\alpha \gtrsim 0$) and elliptical radio galaxies in which the radio emission is concentrated in a compact core. The nuclear radio source in Henize 2-10 has a spectral index ($\alpha \approx -0.4$) similar to Seyfert nuclei that are known to have reduced flux densities on parsec scales. Therefore, we do not consider the non-detection of the nuclear radio source at very high resolution to be incompatible with the presence of an active galactic nucleus in Henize 2-10.

We conclude that an actively accreting massive black hole is the most feasible explanation for the nuclear source in Henize 2-10. The compact radio and hard X-ray luminosities are consistent with the observed correlation for active galactic nuclei and we therefore estimate the mass of the black hole in Henize 2-10 using the so-called “fundamental plane of black hole activity”²⁵. This empirical relationship relating black-hole mass to the emitted compact radio and hard X-ray luminosities, spanning nine orders of magnitude in black-hole mass, is given by the equation $\log L_R = 0.60 \log L_X + 0.78 \log M + 7.33$, where L_R is the radio luminosity at 5 GHz in erg s^{-1} , L_X is the 2–10 keV X-ray luminosity in erg s^{-1} , and M is the mass of the black hole in solar masses, M_\odot . Using the observed radio luminosity of $7.4 \times 10^{35} \text{ erg s}^{-1}$ at 4.9 GHz and the X-ray luminosity of $2.7 \times 10^{39} \text{ erg s}^{-1}$ in the 2–10 keV band, we calculate $\log(M/M_\odot) = 6.3 \pm 1.1$ for the black hole in Henize 2-10. The region in which the gravitational influence of a one-million-solar-mass black hole dominates that of the host galaxy subtends a very small angle on the sky at the distance of Henize 2-10 (< 1 arcsecond for velocity dispersions $> 10 \text{ km s}^{-1}$). Thus, it is not surprising that kinematic studies of Henize 2-10 have not previously revealed the presence of the black hole at its centre.

Few dwarf galaxies are currently known to host massive black holes^{26,27}; however, the discovery of an active nucleus in Henize 2-10 opens up a new realm in which to search for local analogues of primordial black-hole growth (that is, dwarf starburst galaxies). While recent searches^{28,29} have revealed growing numbers of nuclear black holes with masses similar to that in Henize 2-10, the host galaxies of these objects have very different properties from that of Henize 2-10. Most notably, they are not actively forming stars and have regular morphologies of disks and spheroids with well-defined optical nuclei^{29,30}. Moreover, the majority of the black holes detected in these systems are radiating at high fractions of their Eddington limits^{27–29}, suggesting that the black holes are currently undergoing rapid growth. In contrast, the low-luminosity active galactic nucleus in Henize 2-10 is currently radiating significantly below its Eddington limit ($\sim 10^{-4}$ assuming an X-ray bolometric correction of ten; see Supplementary Information), suggesting a different evolutionary state.

The results presented here have broad implications for our understanding of the evolution of galaxies and their central black holes. The concurrent black-hole growth and extreme starburst in Henize 2-10 probably resembles the conditions in low-mass, high-redshift galaxies during the early phases of galaxy assembly when interactions and mergers were common. Indeed, Henize 2-10 shows signs of having undergone an interaction, including tidal-tail-like features in both its gaseous⁷ and stellar distributions (Fig. 1). Additionally, it is intriguing that the massive black hole in Henize 2-10 does not appear to be associated with a bulge, a nuclear star cluster or any other well-defined nucleus. This unusual property may reflect an early phase of black-hole growth and galaxy evolution that has not been previously observed. If so, this implies that primordial seed black holes could have pre-dated their eventual dwellings, thereby constraining theories for the formation mechanisms of massive black holes and galaxies.

Received 1 September; accepted 30 November 2010.

Published online 9 January 2011.

- Magorrian, J. *et al.* The demography of massive dark objects in galaxy centers. *Astron. J.* **115**, 2285–2305 (1998).
- Ghez, A. M. *et al.* Stellar orbits around the galactic center black hole. *Astrophys. J.* **620**, 744–757 (2005).
- Volonteri, M. Formation of supermassive black holes. *Astron. Astrophys. Rev.* **18**, 279–315 (2010).
- Heckman, T. The co-evolution of galaxies and black holes: current status and future prospects. *IAU Symp.* **267**, 3–14 (2010).
- Allen, D. A., Wright, A. E. & Goss, W. M. The dwarf emission galaxy He2–10. *Mon. Not. R. Astron. Soc.* **177**, 91–97 (1976).
- Johnson, K. E., Leitherer, C., Vacca, W. D. & Conti, P. S. Hubble Space Telescope observations of HE 2–10: outflows and young super-star clusters. *Astron. J.* **120**, 1273–1288 (2000).

- Kobulnicky, H. A., Dickey, J. M., Sargent, A. I., Hogg, D. E. & Conti, P. S. Aperture synthesis observations of molecular and atomic gas in the Wolf-Rayet starburst galaxy Henize 2–10. *Astron. J.* **110**, 116–130 (1995).
- Méndez, D. I. *et al.* A complex bipolar outflow in the Wolf-Rayet BCDG He 2–10. *Astron. Astrophys.* **349**, 801–811 (1999).
- Johnson, K. E. & Kobulnicky, H. A. The spectral energy distributions of infant super-star clusters in Henize 2–10 from 7 millimeters to 6 centimeters. *Astrophys. J.* **597**, 923–928 (2003).
- Kobulnicky, H. A. & Martin, C. L. The diffuse and compact X-ray components of the starburst galaxy Henize 2–10. *Astrophys. J.* **718**, 724–738 (2010).
- Engelbracht, C. W. *et al.* Metallicity effects on mid-infrared colors and the 8 μm PAH emission in galaxies. *Astrophys. J.* **628**, L29–L32 (2005).
- Calzetti, D. *et al.* The calibration of mid-infrared star formation rate indicators. *Astrophys. J.* **666**, 870–895 (2007).
- Whitney, B. A. *et al.* Spitzer Sage Survey of the Large Magellanic Cloud. III. Star formation and ~ 1000 new candidate young stellar objects. *Astron. J.* **136**, 18–43 (2008).
- Skrutskie, M. F. *et al.* The Two Micron All Sky Survey (2MASS). *Astron. J.* **131**, 1163–1183 (2006).
- Bell, E. F., McIntosh, D. H., Katz, N. & Weinberg, M. D. The optical and near-infrared properties of galaxies. I. Luminosity and stellar mass functions. *Astrophys. J.* **149** (Supp.), 289–312 (2003).
- van der Marel, R. P., Roeland, P., Alves, D. R., Hardy, E. & Suntzeff, N. B. New understanding of Large Magellanic Cloud structure, dynamics, and orbit from carbon star kinematics. *Astron. J.* **124**, 2639–2663 (2002).
- Kim, S. *et al.* Aperture synthesis mosaic of the Large Magellanic Cloud. *Astrophys. J.* **503**, 674–688 (1998).
- Reines, A. E., Johnson, K. E. & Goss, W. M. Emerging massive star clusters revealed: high-resolution imaging of NGC 4449 from the radio to the ultraviolet. *Astron. J.* **135**, 2222–2239 (2008).
- Reines, A. E., Johnson, K. E. & Hunt, L. K. A new view of the super star clusters in the low-metallicity galaxy SBS 0335–052. *Astron. J.* **136**, 1415–1426 (2008).
- Reines, A. E., Nidever, D. L., Whelan, D. G. & Johnson, K. E. The importance of nebular continuum and line emission in observations of young massive star clusters. *Astrophys. J.* **708**, 26–37 (2010).
- Ho, L. C. Nuclear activity in nearby galaxies. *Annu. Rev. Astron. Astrophys.* **46**, 475–539 (2008).
- Henry, A. L., Turner, J. L., Beck, S. C., Crosthwaite, L. P. & Meier, D. S. Brackett lines from the super star cluster nebulae in HE 2–10. *Astron. J.* **133**, 757–767 (2007).
- Ulvestad, J. S., Johnson, K. E. & Neff, S. G. A VLBI search for radio supernovae in Wolf-Rayet Galaxies. *Astron. J.* **133**, 1868–1873 (2007).
- Orienti, M. & Prieto, M. A. Radio structures of the nuclei of nearby Seyfert galaxies and the nature of the missing diffuse emission. *Mon. Not. R. Astron. Soc.* **401**, 2599–2610 (2010).
- Merloni, A., Heinz, S. & di Matteo, T. A fundamental plane of black hole activity. *Mon. Not. R. Astron. Soc.* **345**, 1057–1076 (2003).
- Filippenko, A. V. & Sargent, W. L. W. Discovery of an extremely low luminosity Seyfert nucleus in the dwarf galaxy NGC 4395. *Astrophys. J.* **342**, L11–L14 (1989).
- Barth, A. J., Ho, L. C., Rutledge, R. E. & Sargent, W. L. W. POX 52: a dwarf Seyfert 1 galaxy with an intermediate-mass black hole. *Astrophys. J.* **607**, 90–102 (2004).
- Greene, J. E. & Ho, L. C. A new sample of low-mass black holes in active galaxies. *Astrophys. J.* **670**, 92–104 (2007).
- Barth, A. J., Greene, J. E. & Ho, L. C. Low-mass Seyfert 2 galaxies in the Sloan Digital Sky Survey. *Astron. J.* **136**, 1179–1200 (2008).
- Greene, J. E., Ho, L. C. & Barth, A. J. Black holes in pseudobulges and spheroidals: a change in the black hole-bulge scaling relations at low mass. *Astrophys. J.* **688**, 159–179 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.E.R. is grateful for many discussions on this work, in particular with M. Whittle, J. Ulvestad, M. Goss, S. Kannappan, J. Greene, R. Hickox, A. Evens, R. O’Connell, R. Chevalier, A. Seth, E. Gallo, S. Ransom, L. Hunt and J. Simon. A.E.R. acknowledges support from a NASA Earth and Space Science Fellowship, and the University of Virginia through a Governor’s Fellowship and a Dissertation Acceleration Fellowship. G.R.S. acknowledges support for this work by NASA through the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA. K.E.J. acknowledges support from the NSF through a CAREER award and the David and Lucile Packard Foundation through a Packard Fellowship. Support was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This research has made use of data obtained from the Hubble Space Telescope and Chandra X-ray Observatory Data Archives.

Author Contributions A.E.R. reduced the HST/NICMOS images, synthesized the multi-wavelength data, and led the analysis, interpretation, and writing of the paper. G.R.S. analysed the Chandra data, and helped with the interpretation and writing of the paper. K.E.J. led the HST/NICMOS and VLA proposals. K.E.J. and C.L.B. reduced and analysed the VLA data. All authors discussed the results and presentation of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.E.R. (areines@virginia.edu).